

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:50:12 ; Search time 64.04 Seconds

(without alignments)  
43.513 Million cell updates/sec

Title: US-09-730-379E-4

Perfect score: 155

Sequence: 1 IGSVRLPLRKGEVLPLPEANFPSPPLP 29

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: PIR.71:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	100.0	525	1 KGHUGH	histidine-rich gly
2	120	77.4	445	2 A60488	histidine-rich gly
3	66.5	42.9	360	2 S09552	nodulin - soybean
4	62	40.0	891	2 H75507	hypothetical prote
5	58.5	37.7	117	2 C71886	hypothetical prote
6	58.5	37.7	1154	2 F71856	hypothetical prote
7	57.5	37.1	162	2 T07173	hypothetical prote
8	54	34.8	428	2 B64081	lucose permease ho
9	54	34.8	624	2 JC5471	regulatory protein
10	53.5	34.5	405	2 T32274	hypothetical prote
11	53.5	34.5	533	2 B56110	tyrosine phosphor
12	53	34.2	599	2 T10798	phosphorin-S - Vo
13	53	34.2	1611	2 T38236	hypothetical prote
14	51	32.9	377	2 T28877	hypothetical prote
15	51	32.9	511	1 B64850	probable virulence
16	51	32.9	511	2 G90809	probable virulence
17	51	32.9	511	2 C85669	probable virulence
18	51	32.9	534	2 T30268	oligo-1,6-glucosid
19	50.5	32.6	461	2 JC4302	tumor necrosis fac
20	50	32.3	497	2 AC0639	virulence factor M
21	50	32.3	516	2 E70908	hypothetical prote
22	50	32.3	524	1 E40271	virulence factor m
23	50	32.3	694	2 A12492	hypothetical prote
24	50	32.3	796	2 T34805	hypothetical prote
25	50	32.3	1015	2 C84918	probable ATP-depen
26	50	32.3	1016	2 H71460	probable outer mem
27	50	32.3	1776	2 G86280	protein T5E21.13 l
28	49.5	31.9	332	2 G96840	hypothetical prote
29	49.5	31.9	984	2 JN0658	restriction endonu

30	49	31.6	325	2 A72724	hypothetical prote
31	49	31.6	361	2 C83350	probable transcrip
32	49	31.6	418	2 F64473	D-alanine transami
33	49	31.6	583	1 US0673	neopullulanase (EC
34	49	31.6	609	2 T38656	probable RNA-bind
35	49	31.6	3149	1 Q0BE8	BpL1 protein - hu
36	48.5	31.3	74	2 T17834	hypothetical prote
37	48.5	31.3	1036	2 B83466	probable RND efflu
38	48	31.0	91	2 S02769	gag 75k protein pr
39	48	31.0	191	2 AD0339	conserved hypothet
40	48	31.0	288	2 H87642	pilin-related prot
41	48	31.0	336	2 B97630	BH0731 conserved h
42	48	31.0	336	2 AD2853	MoxR family protei
43	48	31.0	370	2 T49945	peritaxin-like prot
44	48	31.0	506	2 S13720	coat protein - ara
45	48	31.0	594	2 S50611	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

KGHUGH histidine-rich glycoprotein precursor - human

N:Alternate names: HRG

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #text\_change 16-Jun-2000

C:Accession: A01287; S29669

R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A:Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu

A:Reference number: A01287; MUID:86216149

A:Accession: A01287

A:Molecule type: mRNA

A:Residues: 1-525 <KOI>

A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514

R:Hennis, B.; Havelaar, A.; Kluff, C.

submitted to the EMBL Data Library, October 1991

A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly

A:Reference number: S29669

A:Accession: S29669

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 214-247 <HFN>

A:Cross-references: EMBL:Z17218; NID:g32453; PIDN:CAA78925.1; PID:g32454

C:Comment: Although its physiological function is not yet known, HRG does bind heme,

din, and the lysine-binding site of plasminogen. On the basis of its homology with H

lood coagulation cascade.

C:Comment: The amino half of this protein is homologous to the first two cystatin-lik

could not have inhibitory activity.

C:Comment: In addition to having a high histidine and proline content, this protein f

e-rich region.

C:Genetics:

A:Gene: GDB:HRG

A:Cross-references: GDB:120055; OMIM:142640

A:Map position: 3q27-3q27

C:Superfamily: histidine-rich glycoprotein; cystatin homology

C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F:19-131/Domain: cystatin homology <CY1>

F:140-246/Domain: cystatin homology <CY2>

F:276-321/Region: proline-rich

F:348-437/Region: histidine-rich

F:351-497/Region: proline-rich

F:63.125,344,345/Binding site: carbohydrate (asn) (covalent) #status predicted

F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 100.0%; Score 155; DB 1; Length 525;

Best Local Similarity 100.0%; Pred. No. 1.5e-13;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGSVRLPLRKGEVLPPEANPSPPLP 29  
 DB 455 IGSVRLPLRKGEVLPPEANPSPPLP 483

## RESULT 2

histidine-rich glycoprotein - bovine (fragments)  
 N:Alternate names: autorosette inhibition factor  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 19-Mar-1993 #sequence\_revision 23-Mar-1995 #text\_change 07-Jul-1995  
 C:Accession: S35687; J02196; A60488  
 R:Scorsen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.  
 FEBS Lett. 328, 285-290, 1993  
 A:Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly  
 A:Reference number: S35687; M0ID:93351678  
 A:Accession: S35687  
 A:Molecule type: protein  
 A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>  
 A:Note: 355-Gln and 368-Tyr were also found  
 R:Halaker, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.  
 Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
 A:Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
 A:Reference number: J02196; M0ID:94220160  
 A:Accession: J02196  
 A:Molecule type: protein  
 A:Residues: 1-23;35-54,'VK',57-101,'R';'TVGEYEG',116,'N',118,'R',120-136;137-145;150-20  
 A:Experimental source: Plasma  
 R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halaker, T.  
 Thromb. Res. 60, 385-396, 1990  
 A:Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
 A:Reference number: A60488; M0ID:91196010  
 A:Accession: A60488  
 A:Molecule type: protein  
 A:Residues: 1-6,'X',8-15 <VES>  
 C:Comment: This protein is a single-chained plasma protein which participates in transgl  
 C:Superfamily: histidine-rich glycoprotein; cystatin homology  
 C:Keywords: glycoprotein; plasma  
 F:2-113/Domain: cystatin homology <CY1>  
 F:122-207/Domain: cystatin homology (fragments) <CY2>  
 F:7-424,60-71,87-108,165-146,180-203,258-288/Disulfide bonds: #status experimental  
 F:74,107,164,266/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 77.4%; Score 120; DB 2; Length 445;  
 Best Local Similarity 78.6%; Pred. No. 8,6e-09;  
 Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 GSVRLPLRKGEVLPPEANPSPPLP 29  
 DB 376 GYHRLPLRKGEVLPPEANPSPPLP 403

## RESULT 3

S09552  
 modulin - soybean  
 C:Species: Glycine max (soybean)  
 C:Date: 02-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 09-Sep-1997  
 C:Accession: S09552  
 R:Sengupta-Gopalan, C.; Pitas, J.W.; Thompson, D.V.; Hoffman, L.M.  
 Mol. Gen. Genet. 203, 410-420, 1986  
 A:Title: Expression of host genes during root nodule development in soybeans.  
 A:Reference number: S07315  
 A:Accession: S09552  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-360 <SEN>  
 A:Cross-references: EMBL:X03979; NID:g18717; PID:g18718

Query Match 42.9%; Score 66.5; DB 2; Length 360;  
 Best Local Similarity 40.6%; Pred. No. 0.16;  
 Matches 13; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 1 IGSVRLPLRKGEVLPPEANPSPPLP 29  
 DB 142 LGKVLPPLRPBPGLKILPFGTNIVPFP 173

## RESULT 4

H75507  
 hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: H75507  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; M0ID:20036896  
 A:Accession: H75507  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-891 <WHI>  
 A:Cross-references: GB:AE001911; GB:AE000513; NID:g6458217; PIDN:AAF01012.1; PID:g645  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0530  
 A:Map position: 1

Query Match 40.0%; Score 62; DB 2; Length 891;  
 Best Local Similarity 56.5%; Pred. No. 1.9;  
 Matches 13; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 7 LPPRLRKGEVLPPEANPSPPLP 29  
 DB 366 LPPVAMGEVAPLPVPEVPPPLP 388

## RESULT 5

h71886  
 hypothetical protein jhp0814 - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C:Accession: C71886  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.  
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
 A:Reference number: A71800; M0ID:99120557  
 A:Accession: C71886  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-117 <ARN>  
 A:Cross-references: GB:AE001510; GB:AE001439; NID:g4155371; PIDN:AAD06384.1; PID:g415  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp0814

Query Match 37.7%; Score 58.5; DB 2; Length 117;  
 Best Local Similarity 40.5%; Pred. No. 0.58;  
 Matches 15; Conservative 3; Mismatches 8; Indels 11; Gaps 2;

QY 3 SVYRLPLRKGEVLPPEANPSPPLP 28  
 DB 52 SLVTSPTITGERTPLNSKNKPDPTTNAIKTQEPPLP 88

## RESULT 6

F71856  
 hypothetical protein jhp1044 - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori

C: Superfamily: *Caenorhabditis elegans* hypothetical protein ZC308.2





A:Gene: mv1N  
C:Superfamily: mv1N protein  
C:Keywords: transmembrane protein; virulence  
F:90-106/Domain: transmembrane #status predicted <TM1>  
F:135-151/Domain: transmembrane #status predicted <TM2>  
F:161-177/Domain: transmembrane #status predicted <TM3>  
F:186-202/Domain: transmembrane #status predicted <TM4>  
F:237-253/Domain: transmembrane #status predicted <TM5>  
F:276-292/Domain: transmembrane #status predicted <TM6>  
F:313-329/Domain: transmembrane #status predicted <TM7>  
F:354-370/Domain: transmembrane #status predicted <TM8>  
F:386-402/Domain: transmembrane #status predicted <TM9>  
F:409-425/Domain: transmembrane #status predicted <TM10>  
F:444-460/Domain: transmembrane #status predicted <TM11>  
F:483-499/Domain: transmembrane #status predicted <TM12>

Query Match 32.9%; Score 51; DB 1; Length 511;  
Best Local Similarity 55.0%; Pred. No. 33;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 4 VYRLPLKKEVLPPEANF 23  
II:II I:I :I II II  
Db 202 VYQLPLKKIGMLVLPRIKF 221

Search completed: May 24, 2002, 16:50:13  
Job time: 318 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:49:00 ; Search time 49.37 Seconds  
(without alignments)  
14.348 Million cell updates/sec

Title: US-09-730-379E-4

Perfect score: 155

Sequence: 1 IGSVYRLPLRKGEVLPLPEANFPSEPLP 29

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PT0S.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	31.6	481	3	US-08-787-091-2
2	47.5	30.6	586	3	US-08-459-953A-10
3	47	30.3	168	4	US-08-818-112-100
4	47	30.3	168	4	US-08-818-111-95
5	47	30.3	168	4	US-09-056-556-100
6	47	30.3	187	4	US-08-818-112-63
7	47	30.3	187	4	US-08-818-111-64
8	47	30.3	187	4	US-09-056-556-63
9	47	30.3	338	4	US-08-722-184-4
10	47	30.3	338	4	US-09-043-937A-8
11	47	30.3	951	4	US-08-816-346-58
12	47	30.3	951	4	US-09-335-411-58
13	47	30.3	952	2	US-08-788-674-5
14	47	30.3	952	3	US-08-816-346-4
15	47	30.3	952	4	US-09-335-411-4
16	46	29.7	445	3	US-08-592-900-2
17	46	29.7	445	4	US-09-252-292C-28
18	46	29.7	446	3	US-08-979-917A-1
19	46	29.7	640	3	US-09-026-343-2
20	46	29.7	986	3	US-08-938-291A-7
21	45.5	29.4	85	4	US-09-247-155-133
22	45	29.0	131	4	US-09-199-637A-197
23	45	29.0	342	4	US-09-043-937A-2
24	45	29.0	349	2	US-08-483-926A-12
25	45	29.0	367	2	US-08-737-045-14
26	45	29.0	367	2	US-08-932-871B-2
27	45	29.0	367	3	US-09-476-919-2

28	45	29.0	367	4	US-08-780-311A-2	Sequence 2, Appl1
29	45	29.0	1255	2	US-09-080-897-4	Sequence 4, Appl1
30	45	29.0	1255	3	US-08-899-595-1	Sequence 1, Appl1
31	45	29.0	1255	4	US-09-323-735-5	Sequence 4, Appl1
32	44.5	28.7	688	4	US-08-973-273-26	Sequence 26, Appl1
33	44.5	28.7	747	4	US-08-973-273-3	Sequence 3, Appl1
34	44	28.4	267	4	US-08-454-928-10	Sequence 10, Appl1
35	44	28.4	289	4	US-09-797-906-4	Sequence 4, Appl1
36	44	28.4	289	4	US-08-797-906-5	Sequence 5, Appl1
37	44	28.4	367	1	US-08-147-949A-2	Sequence 2, Appl1
38	44	28.4	574	4	US-09-276-400-6	Sequence 6, Appl1
39	44	28.4	574	4	US-09-448-076-6	Sequence 6, Appl1
40	44	28.4	1274	4	US-09-095-443-2	Sequence 2, Appl1
41	43.5	28.1	483	3	US-08-833-553-2	Sequence 2, Appl1
42	43.5	28.1	483	4	US-09-078-173A-26	Sequence 26, Appl1
43	43.5	28.1	483	4	US-09-418-222-2	Sequence 2, Appl1
44	43.5	28.1	483	4	US-09-537-357-33	Sequence 33, Appl1
45	43	27.7	69	1	US-08-280-443-17	Sequence 17, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-787-091-2  
; Sequence 2, Application US/08787091  
; Patent No. 6100386  
; GENERAL INFORMATION:  
; APPLICANT: Carphino, Nicholas A.  
; APPLICANT: Kobayashi, Ryuji  
; APPLICANT: Wisniewski, David G.  
; APPLICANT: Strife, Annabel O.C.  
; APPLICANT: Clarkson, Bayard D.  
; TITLE OF INVENTION: A No. 6100386e1 Human Gene/Protein Involved In  
; TITLE OF INVENTION: Chronic Myelogenous Leukemia  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,091  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 60/030,418  
; FILING DATE: 01-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: CSHL96-05PA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 481 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-787-091-2  
Query Match 31.6%; Score 49; DB 3; Length 481;  
Best Local Similarity 40.0%; Pred. No. 21;



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/818,111  
APPLICATION NUMBER: 424  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 95:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 168 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-111-95

Query Match 30.3%; Score 47; DB 4; Length 168;  
Best Local Similarity 41.7%; Pred. No. 13;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 VYRLPLRKGEVLPLEAPNSPFP 27  
DB 34 VYOMQPVFGAPLPDPASAPDVP 57

RESULT 5  
US-09-056-556-100  
Sequence 100, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 100:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 168 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-056-556-100

Query Match 30.3%; Score 47; DB 4; Length 168;  
Best Local Similarity 41.7%; Pred. No. 13;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 VYRLPLRKGEVLPLEAPNSPFP 27  
DB 34 VYOMQPVFGAPLPDPASAPDVP 57

RESULT 6  
US-08-818-112-63  
Sequence 63, Application US/08818112  
Patent No. 6290969  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twartzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-63

Query Match 30.3%; Score 47; DB 4; Length 187;  
Best Local Similarity 41.7%; Pred. No. 14;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 VYRLPLRKGEVLPLEAPNSPFP 27  
DB 53 VYOMQPVFGAPLPDPASAPDVP 76

RESULT 7  
US-08-818-111-64  
Sequence 64, Application US/08818111  
Patent No. 6338852  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Nelo, Antonia  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedralik, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
TUBERCULOSIS  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,111  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ. ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-111-64

Query Match 30.3%; Score 47; DB 4; Length 187;  
Best Local Similarity 41.7%; Pred. No. 14;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 VYRLPLRKGEVLPPEANFSPF 27  
DB 53 VYQMOPVFGAPLPDPSADVP 76

RESULT 8  
US-09-056-556-63  
Sequence 63, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
TREATMENT OF  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ. ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-056-556-63

Query Match 30.3%; Score 47; DB 4; Length 187;  
Best Local Similarity 41.7%; Pred. No. 14;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 VYRLPLRKGEVLPPEANFSPF 27  
DB 53 VYQMOPVFGAPLPDPSADVP 76

RESULT 9  
US-08-722-184-4  
Sequence 4, Application US/08722184  
Patent No. 6015943  
GENERAL INFORMATION:  
APPLICANT: Boudet, Alain  
APPLICANT: Pettegnati, Jacqueline  
APPLICANT: Goffner, Deborah  
APPLICANT: Halpin, Claire  
APPLICANT: O'Connell, Ann  
APPLICANT: Ramestant, Michel  
APPLICANT: Boerjan, Wout  
APPLICANT: Lepie, Jean-Charles  
TITLE OF INVENTION: DNA Sequences Coding for a Cinnamoyl CoA  
TRANSFERASE AND THEIR USE IN REGULATION OF PLANT LIGNIN  
CONCENTRATIONS  
TITLE OF INVENTION: Concentrations  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SPENCER & FRANK  
STREET: 1100 New York Ave, N.W., Suite 300E  
CITY: Washington, D.C.  
COUNTRY: US  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,184  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94 04246  
FILING DATE: 11-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Calveit, Frederick C  
REGISTRATION NUMBER: 28,557  
REFERENCE/DOCKET NUMBER: GROBO 7002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)414-4000  
TELEFAX: (202)414-4040  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-722-184-4

Query Match 30.3%; Score 47; DB 3; Length 338;  
Best Local Similarity 40.0%; Pred. No. 28;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 10 LKRGVLPLEPNFSPFLP 29  
| : ||| : : | : | : |  
Db 257 LHRGEVETLAKFFPEYPIP 276

RESULT 10  
US-09-043-937A-8

Sequence 8, Application US/09043937A  
Patent No. 6211432  
GENERAL INFORMATION:

APPLICANT: BOUDET, ALAIN-MICHEL  
PICHON, MAGALIE  
GRIMA-PETENATI, JACQUELINE  
BECKERT, MICHEL  
GAMAS, PASCAL  
BRIAT, JEAN-FRANCOIS

TITLE OF INVENTION: DNA SEQUENCES CODING FOR CINNAMOYL-COA  
REDUCTASE, AND APPLICATIONS THEREOF IN THE CONTROL OF  
LIGNIN CONTENTS IN PLANTS

NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE, P.C.  
STREET: 110 NORTH GLEBE ROAD, 8TH FLOOR  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/043,937A  
FILING DATE: 24-JUL-1998  
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR96/01544  
FILING DATE: 03-OCT-1996  
APPLICATION NUMBER: FR 95.11623  
FILING DATE: 03-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36.663  
REFERENCE/DOCKET NUMBER: 1487-20

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-043-937A-8

Query Match 30.3%; Score 47; DB 4; Length 338;  
Best Local Similarity 40.0%; Pred. No. 28;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 10 LKRGVLPLEPNFSPFLP 29  
| : ||| : : | : | : |  
Db 257 LHRGEVETLAKFFPEYPIP 276

RESULT 11

US-08-816-346-58  
Sequence 58, Application US/08816346  
Patent No. 6127525  
GENERAL INFORMATION:

APPLICANT: Crystal, Ronald G.  
APPLICANT: Falck-Pedersen, Erik  
APPLICANT: Gall, Jason  
APPLICANT: Kovesdi, Imre

TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
METHODS OF USING SAME  
TITLE OF INVENTION: METHODS OF USING SAME  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
STREET: TWO PRUDENTIAL PLAZA - 4900  
CITY: CHICAGO  
STATE: ILLINOIS

COUNTRY: USA  
ZIP: 60601-6780

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,346  
FILING DATE:

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 67167  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/616-5600  
TELEFAX: 312/616-5700  
TELEX: 25-3533

INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 951 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-816-346-58

Query Match 30.3%; Score 47; DB 3; Length 951;  
Best Local Similarity 45.8%; Pred. No. 91;  
Matches 11; Conservative 4; Mismatches 5; Indels 4; Gaps 2;

OY 5 YRLPRLKGEVLPLEPNFSPFL 28  
| : ||| : : | : | : |

Db 830 YLAPYMRGQAVP--ANFP-YPL 849

RESULT 12

US-09-335-411-58  
Sequence 58, Application US/09335411  
Patent No. 6153435  
GENERAL INFORMATION:

APPLICANT: Crystal, Ronald G.  
APPLICANT: Falck-Pedersen, Erik  
APPLICANT: Gall, Jason  
APPLICANT: Kovesdi, Imre

TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
METHODS OF USING SAME  
TITLE OF INVENTION: METHODS OF USING SAME  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
STREET: TWO PRUDENTIAL PLAZA - 4900  
CITY: CHICAGO  
STATE: ILLINOIS





Best Local Similarity 45.8%; Pred. No. 91;  
Matches 11; Conservative 4; Mismatches 5; Indels 4; Gaps 2;

QY 5 YRLPPLRKGEVLPLEAPNFPSPPL 28

Db 831 YLAPTMREGQAYP--ANFP-YPL 850

## RESULT 15

US-09-335-411-4  
; Sequence 4, Application US/09335411  
; Patent No. 6153435  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; TITLE OF INVENTION: METHODS OF USING SAME  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/335,411  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/816,346  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 952 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 951..952  
; OTHER INFORMATION: /note="Xaa can be either Gln, His, or Thr"  
US-09-335-411-4

Query Match 30.3%; Score 47; DB 4; Length 952;  
Best Local Similarity 45.8%; Pred. No. 91;  
Matches 11; Conservative 4; Mismatches 5; Indels 4; Gaps 2;

QY 5 YRLPPLRKGEVLPLEAPNFPSPPL 28

Db 831 YLAPTMREGQAYP--ANFP-YPL 850

Search completed: May 24, 2002, 16:49:01  
Job time: 286 sec

**THIS PAGE BLANK (USPTO)**



PT biodiversity -  
 XX Claim 20; SEQ ID No 57599; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 205 AA:  
 Query Match 38.4%; Score 59.5; DB 22; Length 205;  
 Best Local Similarity 53.6%; Pred. No. 1.5;  
 Matches 15; Conservative 2; Mismatches 10; Indels 1; Gaps 1;  
 OY 2 GSVYRLPPLRKGEVLPLPEANPSPPLP 29  
 II: I I I I : I I I I I I I I  
 Db 69 gsl-trpplpqccclrlprvgfrffplp 95  
 RESULT 2  
 AAO10033 standard; Protein; 130 AA.  
 ID AAO10033:  
 AC AAO10033:  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 23925.  
 XX  
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-514838/56.  
 DR N-PSDB; AAI89964.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -  
 XX Claim 20; SEQ ID No 23925; 1399pp + Sequence Listing; English.  
 PS  
 XX  
 CC The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activating/inhibiting activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 130 AA:  
 Query Match 36.5%; Score 56.5; DB 22; Length 130;  
 Best Local Similarity 48.1%; Pred. No. 2.4;  
 Matches 13; Conservative 1; Mismatches 8; Indels 5; Gaps 1;  
 OY 8 PPLRKGEVLPLPEANPSPF-----PLP 29  
 II I I I I : I I I I I I I I  
 Db 9 pprirkgtpppprkllfpffkppplp 35  
 RESULT 3  
 ABG24782 standard; Protein; 88 AA.  
 ID ABG24782:  
 AC ABG24782:  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #24773.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS88969.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 55141; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The



CC activity. The invention also includes antibodies specific for the  
CC protein or polynucleotide sequences. The lung cancer associated  
CC polynucleotide sequences may be used for detection of lung cancer,  
CC chromosome identification, as chromosome markers, and for numerous other  
CC diagnostic or research purposes. The proteins may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders. The proteins may also be used in the treatment of wounds and  
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
CC peptide AAB58549 are used in the course of the invention for the  
CC identification and characterisation of the polynucleotide and protein  
CC sequences.

XX  
SQ Sequence 95 AA;

Query Match 34.5%; Score 53.5; DB 21; Length 95;  
Best Local Similarity 54.2%; Pred. No. 4.4; 7; Indels 1; Gaps 1;  
Matches 13; Conservative 3; Mismatches

QY 7 LPLRLKGEVLPPEANPSPF-LP 29  
| | | | | : | | | | |  
Db 17 lplrlkgevrvspatpafpslp 40

RESULT 6  
AAO12245 AAO12245 standard; Protein; 124 AA.  
XX  
AC AAO12245;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 26137.  
XX  
KM Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
KM nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-0504927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AAI92176.  
XX  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 26137; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA19941-AA193841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX  
SQ Sequence 124 AA;

Query Match 34.5%; Score 53.5; DB 22; Length 124;  
Best Local Similarity 48.1%; Pred. No. 6;  
Matches 13; Conservative 3; Mismatches 10; Indels 1; Gaps 1;

QY 4 VYRLPLRKGEVLPPEANPSPF-PLP 29  
| : | | | : | | | | : | | |  
Db 2 vflplpkklgqdlpppalrflavsp 28

RESULT 7  
AAG22558 AAG22558 standard; Protein; 113 AA.  
XX  
AC AAG22558;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 25533.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence; corn.  
XX  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129645.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142134.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 27-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 15-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155569.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

# Query Match

Best Local Similarity 32.9%; Score 51; DB 21; Length 113;

Pred. No. 12;

Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 8 PPLRKGEVLPPEANFSPFLP 29

DB 6 pplitrxrvrpyrpsfpfp 27

## RESULT 8

ABB61862 ID ABB61862 standard; Protein; 569 AA.

AC ABB61862;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 12378.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PsDB; ABL05965.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Disclosure; SEQ ID NO 12378; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABBS7737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at fcp.wipo.int/pub/published\_pct\_sequences.

XX

XX Sequence 569 AA;

Query Match 32.9%; Score 51; DB 22; Length 569;

Best Local Similarity 47.6%; Pred. No. 73;

Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 GSVYRLPPLRKGEVLPPEAN 22

DB 458 gsfir1ppgaasaapipe 478

## RESULT 9

AAG81637 ID AAG81637 standard; Protein; 412 AA.

AC AAG81637;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:368.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;

XX vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAX ) GLAXO GROUP LTD.

PI kimmerly wd;

DR WPI; 2001-316495/33.

DR N-PsDB; AAB52487.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

PT useful for vaccinating against infections, e.g. endocarditis -

XX Claim 18; Page 140; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. epidermidis genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to

CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,

CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 412 AA;

QY 3 SVYRLPPLRKGEVLPPEANFSPFL 28

DB 332 amyrlrk-regkdpkpfk1plypi 356

## RESULT 10

AAB65634 ID AAB65634 standard; Protein; 814 AA.

AC AAB65634;

XX



DT 27-MAR-2001 (first entry)  
XX Novel protein kinase, SEQ ID NO: 161.  
DE  
XX  
XX Human: mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
KW immunosuppressive; cardiact; renal; antiinflammatory; antiasthmatic;  
KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
KW immune disorder; cardiovascular disease; neurodegenerative disease;  
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
KW inflammatory pelvic disease; multiple sclerosis; psoriasis.  
XX  
XX Homo sapiens.  
OS  
XX MO200073469-A2.  
PN  
XX  
XX 07-DEC-2000.  
PD  
XX  
XX 26-MAY-2000; 2000MO-US14842.  
PF  
XX  
XX 28-MAY-1999; 99US-0136503.  
PR  
XX  
XX (SUGC-) SUGEN INC.  
PA  
XX  
XX Plowman GD, Martinez R, Whyte D, Sudersanam S;  
PI  
XX  
XX WPI: 2001-032161/04.  
DR  
XX  
XX N-PSDB; AAF44661.  
DR  
XX  
XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
PT treating immune-related diseases and disorders, cardiovascular disease,  
PT neurodegenerative diseases and/or cancers -  
XX  
XX  
XX Claim 10: Fig 1; 310pp; English.  
PS  
XX  
XX The present sequence is a novel protein kinase. The novel protein kinases  
CC and the nucleic acids that encode them may be used in the treatment and  
CC diagnosis of diseases associated with inappropriate kinase expression  
CC such as immune-related diseases and disorders, cardiovascular disease,  
CC neurodegenerative diseases and/or cancers. The nucleic acids and  
CC complementary sequences may also be used as DNA probes in diagnostic  
CC assays. The kinase polypeptides may be used as antigens in the production  
CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
CC and kinase antagonists may also be used to down regulate kinase  
CC expression and activity. Diseases related to kinase expression and  
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
CC disorders, complications of organ transplantation, myocardial infarction,  
CC immune disorders, cardiomyopathies, strokes, renal failure,  
CC oxidative-stress related disorders, chronic inflammatory bowel disease,  
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
CC reproductive disorders.  
XX  
XX Sequence 814 AA:  
SQ

Query Match 32.6%; Score 50.5; DB 22; Length 814;  
Best Local Similarity 31.1%; Pred. No. 1.3e+02;  
Matches 14; Conservative 5; Mismatches 9; Indels 17; Gaps 2;

QY 2 GSVYRLPPLR-----KGEVLPLPEANFPS-FPLP 29  
I: |||||:| :| | | |||  
Db 226 gtaerlpprtasglgadpagaavvpsgqdgvpdpagaftpghlplp 270

RESULT 11  
AAV42111  
XX AAV42111 standard; Protein; 819 AA.  
XX  
AC AAV42111;  
XX  
DT 13-DEC-1999 (first entry)  
XX  
DE Human ischaemic heart disease associated protein.

XX  
XX Human: ischaemic heart disease; ischaemia; heart tissue; detection;  
KW ischaemic cardiomyopathy; diagnosis.  
XX  
XX  
XX Homo sapiens.  
OS  
XX MO9949062-A1.  
PN  
XX  
XX 30-SEP-1999.  
PD  
XX  
XX 26-MAR-1999; 99MO-US06662.  
PF  
XX  
XX 26-MAR-1998; 98US-0079377.  
PR  
XX  
XX (GENE-) GENE LOGIC INC.  
PA  
XX  
XX Frashar Y, Fan W;  
PI  
XX  
XX WPI: 1999-580453/49.  
DR  
XX  
XX N-PSDB; AA225100.  
DR  
XX  
XX New isolated nucleic acids associated with ischemic heart tissue, used  
PT to develop products for treating e.g. ischaemic heart disease and  
PT ischaemic cardiomyopathy -  
XX  
XX  
XX Claim 14: page 44-47; 51pp; English.  
PS  
XX  
XX The present sequence represents human ischaemic heart disease associated  
CC protein (I). The (I) nucleotide sequence is up-regulated in ischaemic  
CC heart tissue. The (I) proteins can serve as a target for agents that  
CC can be used to modulate the expression or activity of the protein, e.g.  
CC agents may be identified which modulate biological processes associated  
CC with ischaemic injury to the heart such as chronic ischaemic heart  
CC disease and ischaemic cardiomyopathy. Agents may also be identified  
CC which modulate the biological process associated with recovery to  
CC ischaemic injury to the heart. They can also be used for isolated  
CC binding partners that bind to the protein. The proteins also provide a  
CC novel target for screening of synthetic small molecules and  
CC combinatorial or naturally occurring compound libraries to discover  
CC novel therapeutics to regulate heart function. The products can also be  
CC used for detection, diagnosis and production of transgenic animals.  
XX  
XX Sequence 819 AA:  
SQ

Query Match 32.6%; Score 50.5; DB 20; Length 819;  
Best Local Similarity 31.1%; Pred. No. 1.3e+02;  
Matches 14; Conservative 5; Mismatches 9; Indels 17; Gaps 2;

QY 2 GSVYRLPPLR-----KGEVLPLPEANFPS-FPLP 29  
I: |||||:| :| | | |||  
Db 199 gtaerlpprtasglgadpagaavvpsgqdgvpdpagaftpghlplp 243

RESULT 12  
AAW10034  
XX AAW10034 standard; Protein; 875 AA.  
XX  
AC AAW10034;  
XX  
DT 24-SEP-1997 (first entry)  
XX  
DE Beta-xylosidase regulator (xyIR).  
XX  
XX xyID: xIND1 beta-xylosidase; regulator; xyIR; bread; xylose production;  
KW sweetener; xyitol; animal feed.  
XX  
OS Aspergillus niger.  
XX  
XX WO9700964-A1.  
PN  
XX  
XX 09-JAN-1997.  
PD  
XX

PF 24-JUN-1996; 96WO-NL00258.  
 XX  
 PR 23-JUN-1995; 95EP-0201707.  
 XX  
 PA (RIJK-) RIJKS-LANDBOUWHOGESCHOOL.  
 XX  
 PI De Graaf LH, Van Den Broeck HC, Van Pelj NMK, Visser J;  
 XX  
 DR WPI; 1997-087391/08.  
 DR N-P-SDB; AAT70124.  
 XX  
 PT New isolated beta-xylosidase and related gene - used for the prodn.  
 PT of xylose and oligomers suitable for use as sweeteners or for  
 PT improving bread or animal feed  
 PS Disclosure; Page 28-33; 45pp; English.  
 XX  
 CC The xylR protein is encoded by the xlnR structural gene, isolated from  
 CC Aspergillus niger. XylR is an activating regulator protein of the  
 CC xylanolytic pathway. The xylR protein activates beta-xylosidase (BX). BX  
 CC peptides can be used in bread doughs and other bakery products, resulting  
 CC in improved bread characteristics. They can also be used for the  
 CC production of xylose and xylose oligomers from wood and plant wastes and  
 CC spent paper pulp, which xylose and oligomers are suitable as sweeteners.  
 CC They can also be reduced to xylitol, which is also an effective bulk  
 CC sweetener. Host cells, where the BX gene has been disrupted, can be  
 CC used, e.g. in the production of recombinant enzymes and enzyme  
 CC preparations, e.g. to be added to animal feed.  
 XX  
 SO Sequence 875 AA;

Query Match 32.6%; Score 50.5; DB 18; Length 875;  
 Best Local Similarity 54.2%; Pred. No. 1.4e+02;  
 Matches 13; Conservative 1; Mismatches 7; Indels 3; Gaps 1;

QY 8 PPLRKGEVLPPE---ANFSPFL 28  
 II : IIII IIIII I  
 Db 263 PPQSPGWLPDPSpanfisl 286

## RESULT 13

AAW08586 standard; Protein: 875 AA.

AC AAW08586;

DT 26-MAR-1997 (first entry)

DE A. niger xylanase regulator xylR.

KW xylanase regulator; xlnR; xylR; DNA cassette; selectable marker;  
 KM upstream activating sequence.  
 XX

OS Aspergillus niger N400 (CBS 120.49).

XX  
 FH Key Location/Qualifiers  
 FT Domain 55..83  
 FT /label= "zinc finger binding domain  
 FT /note= "zinc binuclear cluster domain contains 6  
 FT cysteine residues coordinating with zinc"  
 FT Region 538..543  
 FT /label= RRRLLW\_motif  
 FT /note= "(Claim 57, page 93)"  
 XX

PN W09700962-A1.

PD 09-JAN-1997.

PF 24-JUN-1996; 96WO-NL00259.

PR 30-AUG-1995; 95EP-0202346.  
 PR 23-JUN-1995; 95EP-0201707.

XX  
 PA (RIJK-) RIJKS-LANDBOUWHOGESCHOOL WAGENINGEN.  
 XX  
 PI De Graaf LH, Van Den Broeck HC, Visser J;  
 XX  
 DR WPI; 1997-087389/08.  
 DR N-P-SDB; AAT50875.  
 XX  
 PT Nucleic acid cassette contg. bidirectional marker gene and inducible  
 PT enhancer or activator - derived from metabolically active gene, used  
 PT in prep. of mutants with specifically increased or reduced  
 PT metabolism  
 PS Claim 48; Page 75-79; 103pp; English.  
 XX  
 CC The A. niger regulator of the xylanolytic pathway, XylR (AAW08586),  
 CC is a DNA binding protein encoded by the xylR gene (AAT50875). The  
 CC targets for the regulator include the genes xlnA, xlnB, xlnC, xlnD  
 CC and axaA. A knockout mutant host cell lacking XylR is free of  
 CC xylanolytic activity and can be used for prodn. of homologous or  
 CC heterologous proteins free of xylanolytic side activities.  
 XX  
 SO Sequence 875 AA;

Query Match 32.6%; Score 50.5; DB 18; Length 875;  
 Best Local Similarity 54.2%; Pred. No. 1.4e+02;  
 Matches 13; Conservative 1; Mismatches 7; Indels 3; Gaps 1;

QY 8 PPLRKGEVLPPE---ANFSPFL 28  
 II : IIII IIIII I  
 Db 263 PPQSPGWLPDPSpanfisl 286

## RESULT 14

ABG20347 standard; Protein: 937 AA.

AC ABG20347;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #20338.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-P-SDB; AAS84534.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 20; SEQ ID No 50706; 103pp; English.

XX The invention relates to isolated polynucleotide (1) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 937 AA:

Query Match 32.6%; Score 50.5; DB 22; Length 937;  
 Best Local Similarity 31.1%; Pred. No. 1.5e+02;  
 Matches 14; Conservative 5; Mismatches 9; Indels 17; Gaps 2;

OY 2 GSYVRLPPLR-----KGEVLPLPEANFPS-FPLP 29  
 | : ||||| : | : | | | | |  
 Db 165 gtaerlprlasg1gadpaqavvpsgvgdypgpaqafpghlplp 209

## RESULT 15

AAB20301 AAB20301 standard; Protein; 119 AA.

AC AAB20301;

DT 29-MAY-2001 (first entry)

DE Arabidopsis apoptosis inducing factor-like protein.

KM Apoptosis inducing factor: AIF; monodehydroascorbate reductase;

KW male sterility; stress response; tissue culture; plant.

OS Arabidopsis thaliana.

OS Arabidopsis thaliana.

PD 22-MAR-2001.

PF 11-SEP-2000; 2000WO-US24859.

PR 13-SEP-1999; 99US-0153737.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Butler KH, Cahoon RE, Klein TW, Mazithulela G;

DR N-PSDB; AAF30430.

DR N-PSDB; AAF30430.

PT New polynucleotide encoding plant apoptosis inducing factor for  
 PT engineering male sterility, altering plant architecture, manipulating  
 PT stress response, and influencing growth of cells and tissues in culture

PS Claim 1(d); Page 47-48; 91pp; English.

CC The present sequence is that of a monodehydroascorbate reductase  
 CC like apoptosis inducing factor (AIF) of Arabidopsis thaliana  
 CC Landsberg erecta. The sequence is predicted from non-full-length

CC cDNA (see AAF30430) of a clone isolated from a fertilised carpel  
 CC cDNA library. It shows 60% amino acid identity to previously known  
 CC monodehydroascorbate-reductase-like AIFs of Oryza sativa, Brassica  
 CC juncea and Arabidopsis thaliana. The invention relates to isolated  
 CC nucleic acids encoding plant AIFs. Such nucleic acids, used in  
 CC sense or antisense constructs, can be used to alter levels of AIFs  
 CC in plant cells, and hence to alter apoptosis and eventually to  
 CC control cell tissue culture growth, facilitate studies of programmed  
 CC cell death in plants, increase the efficiency of gene transfer, help  
 CC provide more stable transformations, engineer male sterility, alter  
 CC plant architecture and manipulate stress response.  
 CC  
 SO Sequence 119 AA:

Query Match 32.3%; Score 50; DB 22; Length 119;  
 Best Local Similarity 50.0%; Pred. No. 18;  
 Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 5 YRLPPLRKGEVLPLPEANFPSF 26  
 | | | | | | | | | | |  
 Db 44 yerpalskgflpeaparipsf 65

Search completed: May 24, 2002, 16:48:05  
 Job time: 245 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:02:53 ; Search time 107.9 Seconds  
(without alignments)  
46.495 Million cell updates/sec

Title: US-09-730-379E-5  
Perfect score: 159  
Sequence: 1 DNQPFQSVSSESCPGKFKSGFPQVSMFPT 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	119	74.8	525	11	Q99PS6	Q99PS6 mus musculus
2	112	70.4	515	11	Q99PS7	Q99PS7 rattus norv
3	111	69.8	525	11	Q9ESB3	Q9ESB3 mus musculus
4	111	69.8	525	11	Q99PS5	Q99PS5 mus musculus
5	104	65.4	510	11	Q9ESB2	Q9ESB2 rattus norv
6	104	65.4	525	11	Q99PS8	Q99PS8 rattus norv
7	83	52.2	546	6	Q98GCU1	Q98GCU1 bos taurus
8	56.5	33.5	174	16	Q9CLK1	Q9CLK1 pasteurella
9	55	34.6	592	5	Q9VTS5	Q9VTS5 drosophila
10	53	33.3	172	5	Q9W5T3	Q9W5T3 drosophila
11	53	33.3	185	5	Q16157	Q16157 drosophila
12	52	32.7	348	10	Q48631	Q48631 prunus armenica
13	52	32.7	654	4	Q9H593	Q9H593 homo sapien
14	52	32.7	726	4	Q9HCB9	Q9HCB9 homo sapien
15	52	32.7	928	4	Q9B208	Q9B208 homo sapien
16	52	32.7	928	4	Q9H9Y8	Q9H9Y8 homo sapien

17	50.5	31.8	203	16	Q9ZSC0	Q9ZSC0 rhizobium m
18	50	31.4	725	11	Q9ESL7	Q9ESL7 mus musculus
19	50	31.4	736	11	Q9ESN0	Q9ESN0 rattus norv
20	50	31.4	1021	5	O15733	O15733 dictyostell
21	50	31.4	1023	11	P70588	P70588 rattus norv
22	50	31.4	1032	11	Q9QX27	Q9QX27 rattus norv
23	49.5	31.1	518	10	Q9SA38	Q9SA38 arabidopsis
24	49.5	31.1	1360	3	Q9PA22	Q9PA22 neurospora
25	49	30.8	166	13	Q9PE41	Q9PE41 gallus gall
26	49	30.8	279	10	Q41543	Q41543 tritlicum ae
27	49	30.8	298	10	Q94G92	Q94G92 tritlicum ae
28	49	30.8	300	5	Q95286	Q95286 leishmania
29	49	30.8	455	10	Q9PR41	Q9PR41 secale cere
30	48	30.2	235	8	Q956C6	Q956C6 inversidens
31	48	30.2	280	10	Q9RW7	Q9RW7 tritlicum ae
32	47.5	29.9	527	10	O64515	O64515 arabidopsis
33	47	29.6	135	5	Q9VAX2	Q9VAX2 drosophila
34	47	29.6	185	4	Q96HA4	Q96HA4 homo sapien
35	47	29.6	198	4	Q9NWV0	Q9NWV0 homo sapien
36	47	29.6	202	5	Q9V611	Q9V611 drosophila
37	47	29.6	312	11	Q99PE2	Q99PE2 mus musculus
38	47	29.6	332	8	Q97RK6	Q97RK6 crassostrea
39	47	29.6	374	5	Q9N516	Q9N516 caenorhabdi
40	47	29.6	486	3	Q05838	Q05838 saccharomyc
41	47	29.6	1018	10	Q9M3F4	Q9M3F4 arabidopsis
42	46.5	29.2	514	10	Q39228	Q39228 arabidopsis
43	46.5	29.2	523	11	Q9QYW0	Q9QYW0 rattus norv
44	46.5	29.2	1274	5	O16568	O16568 caenorhabdi
45	46.5	29.2	2567	13	Q90574	Q90574 gallus gall

## ALIGNMENTS

RESULT	ID	Q99PS6	PRELIMINARY;	PRT;	525 AA.
AC	Q99PS6				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2001	(TREMBLrel. 18, Last annotation update)			
DE	HISTIDINE-RICH GLYCOPROTEIN.				
GN	MMHRG.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SRRAIN-BALB/C; TISSUE=LIVER;				
RA	Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,				
RA	Onashi T., Sato N., Hirata D., Tsuchida N., Koide T.;				
RT	Molecular diversity of mammalian histidine-rich glycoprotein."				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB055897; BAB33094.1; -				
DR	Interpro: IPR000010; Cystatin.				
DR	Pfam: PF00031; Cystatin; 1.				
DR	SMART: SM00043; Cyt 2.				
SO	SEQUENCE	525 AA;	59090 MW;	A83E93A439CFB3AC	CRC64;

Query Match 74.8%; Score 119; DB 11; Length 525;  
Best Local Similarity 76.9%; Pred. No. 2e-10; 3; Indels 0; Gaps 0;  
Matches 20; Conservative 3; Mismatches 3;

QY 3 QPFPQSVSSESCPGKFKSGFPQVSMFPT 28  
|||||:| |||||:|||||:| |  
Db 494 QPFPQSVSSESCPGKFKSGFPQVSMFPT 519

RESULT 2  
Q99PS7 PRELIMINARY; PRT; 515 AA.

AC Q99P57: 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)  
 DT 01-OCT-2001 (TReMBLrel. 18, last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN 2.  
 GN RNRHG2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
 RA Wakahayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB055896; BAB33093.1; -;  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF000031; cystatin; 1.  
 DR SMART; SM00043; Cy\_2  
 SQ SEQUENCE 515 AA; 5805 MW; 7CBEA3A1A3678966 CRC64;

Query Match	70.4%	Score 112;	DB 11;	Length 515;
Best Local Similarity	76.9%	Pred. No. 2.5e+09;		
Matches 20; Conservative	3;	Mismatches	3;	Indels 0; Gaps 0

```
QY      3 QPFPQSVSESCPGKFKSGFPQVSMFF 28
          |||||:|:|||||||:| ||||| ||
Db     484 QPFPQTASKSCSPGKFESKFPQVSNFF 509
```

RESULT	3	
09ESB3		
ID	09ESB3	PRELIMINARY;
AC	09ESB3	PRT; 525 AA.
DT	01-MAR-2001 (Tremblrel. 16, Created)	
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)	
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)	
DE	HISTIDINE-RICH GLYCOPROTEIN.	
GN	HRG.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=129;	
RX	MEDLINE=20307726; PubMed=10849117;	
RT	Hulett M.D., Parish C.R.;	
RT	"Murine histidine-rich glycoprotein: cloning, characterization and	
RT	cellular origin.";	
RL	Immunol. Cell Biol. 78:280-287 (2000).	
DR	EMBL; AF194028; A628416.1; -	
DR	InterPro: IPR000010; Cystatin.	
DR	Pfam: PF00003; cystatin. 1.	
DR	SMART: SM00043; CY; 2.	
QO	SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;	

Query Match	69.8%	Score 111;	DB 11;	Length 525;
Best Local Similarity	73.1%	Pred. No. 3.6e-09;		
Matches 19; Conservative	3;	Mismatches 4;	Indels 0;	Gaps 0;

```
QY      3 QPFPQSVSESCPGKFKSGFPQVSMFF 28
         |||||:| |||||:| |||:| |
Db      494 QPFPQTASRSCPGKFESEFPQISKFF 519
```

RESULT	4
Q99PS5	
ID	Q99PS5
PRELIMINARY;	
PRT;	525 AA

AC 039P55  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DR 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088) .  
GN MHRG.  
OS Mus musculus (Mouse) .  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
OX NCBI\_TaxID=10090;  
[1]  
RN SEQUENCE FROM N.A.  
RA Tsuchida N., Wakabayashi S., Jahnen-Dechent W., Koide T.;  
RT "Structure of mouse histidine-rich glycoprotein gene".  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RA TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;  
RC Strausberg R.;  
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB055898; BAB33095.1; -  
DR EMBL; BC011668; AAH11668.1; -  
DR InterPro; IPRO00010; Cystatin.  
DR Pfam; PF00031; Cystatin, 1.  
DR SMART; SM00043; Cy; 2  
SO SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match	69.8%	Score 111	DB 11	length 525
Best Local Similarity	73.1%	Pred. No. 3.6e-09		
Matches 19	Conservative 3	Mismatches 4	Indels 0	Gaps 0

```

QY      3 QPFPQSVSESCPGKFKSGFPQVSMFF 28
         ||||:| |||||:| |||:| |
Db     494 QPFPQTASRSCPGKFESEFPQISKFF 519

```

RESULT	5			
09ESB2				
ID	09ESB2	PRELIMINARY;	PRT;	510 AA.
AC	09ESB2;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HISTIDINE-RICH GLYCOPROTEIN.			
GN	HRC.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LEWIS;			
RX	MEDLINE=20307726; PubMed=10849117;			
RA	Hulet M.D., Parish C.R.;			
RT	"Murine histidine-rich glycoprotein: cloning, characterization and			
RL	cellular origin.";			
DR	Immunol. Cell Biol. 78:280-287 (2000).			
DR	EMBL: AF194029; AAC8417.1; "			
DR	InterPro: IPR000010; Cystatin.			
DR	Pfam: PF00003; cystatin, 1.			
DR	SMART: SM00043; CY; 2.			
QO	SEQUENCE	510 AA;	57581 MW;	5086E06A2AD58E CRC64;

Query Match	65.4%	Score 104	DB 11	Length 510
Best Local Similarity	69.2%	Pred. No. 4.5e-08		
Matches 18; Conservative	4	Mismatches 4	Indels 0	Gaps 0

QY    3   QPFPQSVSSESCPGKFKSGFPPQVSMFF    28  
       :|||:|||||:|||||  
Db   479 RPFPTASKSCPGKFEGKFPQVSNFF    504



RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Meriklov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclele J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003543; AM222321.1; -;  
 DR FLYBase: FBgn0036240; CG6928.  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; STAS; 1.  
 DR Pfam: PF00916; Sulfate\_transp; 1.  
 SO SEQUENCE 592 AA; 62941 MW; 22000286CDE096F8 CRC64;

Query Match 34.6%; Score 55; DB 5; Length 592;  
 Best Local Similarity 60.0%; Pred. No. 2.9;  
 Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

OY 11 ESCP-----GKFKSGFPQVSM 26  
 ||||| ||||| |||||  
 Db 229 ESCPFLITGKVGKSLPWSL 248

RESULT 10  
 ID 09W573 PRELIMINARY; PRT; 172 AA.  
 AC 09W573;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE CG15848 PROTEIN.  
 GN SCPI OR CG15848.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Burton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bayendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadenhead E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Meriklov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclele J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE002777; AF45357.1; -;  
 DR FLYBase: FBgn0020908; Scpi.  
 DR InterPro: IPR002048; EF-hand.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 SO SEQUENCE 172 AA; 19582 MW; 283381FEE39EA96D CRC64;

Query Match 33.3%; Score 53; DB 5; Length 172;  
 Best Local Similarity 40.7%; Pred. No. 1.6;  
 Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

OY 2 NQEPPOSSESCKPKFKSGFPQVSMF 28  
 ||||| :||| |||||  
 Db 57 NQEPKDAVKTKCVGKKYEEFPQAMRAF 83

RESULT 11  
 ID 016157 PRELIMINARY; PRT; 185 AA.  
 AC 016157;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE CALCIUM-BINDING PROTEIN (FRAGMENT).  
 GN SCPI OR CG15848.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON-R; TISSUE=MUSCLE;  
 RA Kelly L.E., Phillips A.M., Delbridge M., Stewart R.;  
 RL Insect Biochem. Mol. Biol. 0:0-0(1997).  
 CC -1-SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 DR EMBL: AF014951; AB67804.1; -;  
 DR FLYBase: FBgn0020908; Scpi.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand; 3.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_3.  
 KW Calcium-binding.  
 FT NON\_TER 1  
 SO SEQUENCE 185 AA; 21195 MW; 4B01C5BDBEC8E23D CRC64;

Query Match 33.3%; Score 53; DB 5; Length 185;  
 Best Local Similarity 40.7%; Pred. No. 1.7;  
 Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;  
 OY 2 NQEPPOSSESCKPKFKSGFPQVSMF 28  
 ||||| :||| |||||  
 Db 75 NQEPKDAVKTKCVGKKYEEFPQAMRAF 101



RESULT 12  
 048631  
 ID 048631 PRELIMINARY: PRT: 348 AA.  
 AC 048631:  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE EFTYLENE-FORMING-ENZYME-LIKE DIOXYGENASE.  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
 OX NCBI\_TaxID=36596;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERGERON; TISSUE-MESOCARP, EXOCARP;  
 RA Mdeguie-A-Mdeguie D., Gomez R.-M., Ellis-Lycaon B.,  
 RT "Molecular cloning and nucleotide sequence of an EFE-like dioxygenase  
 protein from apricot."  
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U97530; AAB88878.1; -; Fe\_Asc\_oxidored.  
 DR InterPro: IPR002419; Fe\_Asc\_oxidored; 1.  
 DR Pfam: PF00671; Fe\_Asc\_oxidored; 1.  
 KW Dioxygenase.  
 SQ SEQUENCE 348 AA; 39538 MW; 46894AEB5B2AF31A CRC64;

Query Match 32.7%; Score 52; DB 10; Length 348;  
 Best Local Similarity 39.3%; Pred. No. 4.9;  
 Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

OY 1 DNOPPOSVSESCPGKFKSGFPQVSMFP 28  
 DB 306 DIEPFESLVNESTPOLIKKVKNYGIVT 333

RESULT 13  
 09H593  
 ID 09H593 PRELIMINARY: PRT: 654 AA.  
 AC 09H593:  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE BG56G5.1 (NOVEL PROTEIN) (FRAGMENT).  
 GN BG56G5.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallis J.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL116086; CAC10332.1; -;  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR PROSITE: PS00636; DnaJ\_1; UNKNOWN\_1.  
 FT NON-TER  
 DR INTERPRO: IPR001623; DnaJ\_N.  
 DR PROSITE: PS00636; DnaJ\_1; UNKNOWN\_1.  
 SQ SEQUENCE 654 AA; 71397 MW; 5FF8DBA47295ED91 CRC64;

Query Match 32.7%; Score 52; DB 4; Length 654;  
 Best Local Similarity 42.3%; Pred. No. 9.8;  
 Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 3 QPPOSVSESCPGKFKSGFPQVSMFP 28  
 DB 70 QPFLASILEELMGPVSSGFEVRVLF 95

RESULT 14  
 09HCB9  
 ID 09HCB9 PRELIMINARY: PRT: 726 AA.  
 AC 09HCB9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NIBAN.  
 GN NIBAN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE=20546198; PubMed=11011112;  
 RA Matjima S., Kajino K., Fukuda T., Otsuka F., Hino O.;  
 RT "A Novel Gene 'Niban' Upregulated in Renal Carcinogenesis: cloning by  
 the cDNA-Amplified Fragment Length Polymorphism Approach."  
 RL Jpn. J. Cancer Res. 91:869-874(2000).  
 DR EMBL: AB050477; BAB17230.1; -;  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR PROSITE: PS00636; DnaJ\_1; UNKNOWN\_1.  
 SQ SEQUENCE 726 AA; 79855 MW; 3f6F6012BFB6CD1F CRC64;

Query Match 32.7%; Score 52; DB 4; Length 726;  
 Best Local Similarity 42.3%; Pred. No. 11;  
 Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 3 QPPOSVSESCPGKFKSGFPQVSMFP 28  
 DB 142 QPFLASILEELMGPVSSGFEVRVLF 167

RESULT 15  
 09BZ08  
 ID 09BZ08 PRELIMINARY: PRT: 928 AA.  
 AC 09BZ08:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Clorf24.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21218927; PubMed=11318611;  
 RA Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,  
 RA Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,  
 RA Graham C., Baxevas A.D., Klinger K.W., Landes G.M., Trent J.M.,  
 RA Carpen J.D.;  
 RT "Cloning and characterization of 13 novel transcripts and the human  
 RT RGS8 gene from the 1q25 region encompassing the hereditary prostate  
 RT cancer (hpc1) locus."  
 RL Genomics 73:211-222(2001).  
 DR EMBL: AF288391; AAG60611.1; -;  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR PROSITE: PS00636; DnaJ\_1; UNKNOWN\_1.  
 SQ SEQUENCE 928 AA; 103134 MW; 93717F41336488FE CRC64;

Query Match 32.7%; Score 52; DB 4; Length 928;  
 Best Local Similarity 42.3%; Pred. No. 14;  
 Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 3 QPPOSVSESCPGKFKSGFPQVSMFP 28  
 DB 344 QPFLASILEELMGPVSSGFEVRVLF 369

Search completed: May 24, 2002, 17:02:55  
 Job time: 885 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 24, 2002, 16:44:55 ; Search time 64.04 Seconds  
(without alignments)  
112.534 Million cell updates/sec

Title: US-09-730-379E-1  
Perfect score: 424

Sequence: 1 GPRPFHCRDQIGSVYRLPLR.....SESCPGKFKSGPQVSMFF 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	424	100.0	525	1 KGHUGH	histidine-rich gly
2	246	58.0	445	2 A60488	histidine-rich gly
3	77.5	18.3	260	2 S18350	seed storage prote
4	77	18.2	360	2 S09552	nodulin - soybean
5	76.5	18.0	313	2 S66312	G-box binding fact
6	75	17.7	291	1 EEW7G	gamma-gliadin B pr
7	75	17.7	310	2 T06211	C-hordein precursor
8	75	17.7	456	2 T35474	50kD proline-rich
9	74	17.5	707	2 S68858	finger protein - m
10	73.5	17.3	357	2 S18235	omega secalin prec
11	73.5	17.3	538	1 S12570	homeotic protein b
12	73	17.2	347	2 T05737	probable hordein C
13	72.5	17.1	305	2 S08312	gamma-hordein-1 pr
14	72	17.0	251	2 PS0094	gamma-gliadin prec
15	72	17.0	357	2 S18236	omega secalin prec
16	71	16.7	124	2 E84613	hypothetical prote
17	71	16.7	900	2 B87957	protein Y1066D.7
18	71	16.7	948	2 T20801	hypothetical prote
19	70	16.5	1126	2 T20801	hypothetical prote
20	69.5	16.4	1743	2 T26859	hypothetical prote
21	69	16.3	494	1 W0FFBC	homeotic protein b
22	68.5	16.2	162	2 T07173	hypothetical prote
23	68.5	16.2	302	2 J0A153	gamma-gliadin prec
24	68.5	16.2	1119	2 T16720	hypothetical prote
25	68	16.0	1355	2 S40022	spalt protein - fr
26	68	16.0	1745	2 A46431	tight junction-ass
27	67.5	15.9	148	2 JH0788	sheath glycoprotei
28	67	15.8	111	2 E69986	hypothetical prote
29	67	15.8	241	1 RRM218	phosphoprotein P -

30	67	15.8	248	2 T51056	hypothetical prote
31	67	15.8	1142	2 T37455	enmein precursor
32	66.5	15.7	253	2 T17312	hypothetical prote
33	66.5	15.7	481	1 B43674	protein kinase (EC
34	66.5	15.7	542	2 A44358	zyxin - chicken
35	66.5	15.7	1776	2 G86280	protein Y5E21.13 l
36	66.5	15.7	1874	1 J00533	genome polypeptide
37	66	15.6	264	2 S07975	B3-hordein (clone
38	66	15.6	594	2 S50611	hypothetical prote
39	65.5	15.4	445	2 S00256	Krox-20 protein -
40	65.5	15.4	520	2 E97813	WASP, N-WASP, MENA
41	65.5	15.4	533	2 T39025	hypothetical prote
42	65.5	15.4	827	2 T39608	zinc finger transc
43	65.5	15.4	1528	2 S13743	DNA strand transfe
44	65.5	15.4	1607	2 T43212	insulin-like growt
45	65	15.3	782	2 I48746	semaphorin C - mou

## ALIGNMENTS

## RESULT 1

KGHUGH histidine-rich glycoprotein precursor - human

N:Alternate names: HRG

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000

C:Accession: A01287; S29669

R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A:Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu

A:Reference number: A01287; MUID:86216149

A:Accession: A01287

A:Molecule type: mRNA

A:Residues: 1-525 <KOI>

A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514

A:Comment: Although its physiological function is not yet known, HRG does bind heme, din, and the lysine-binding site of plasminogen. On the basis of its homology with HM

lood coagulation cascade.

C:Comment: The amino half of this protein is homologous to the first two cystatin-lik

C:Comment: In addition to having a high histidine and proline content, this protein h

e-rich' region.

C:Genetics:

A:Gene: GDB:HRG

A:Cross-references: GDB:120055; OMIM:142640

A:Map position: 3q27-3q27

C:Superfamily: histidine-rich glycoprotein; cystatin homology

C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F:1-18/Domain: signal sequence #status predicted <SIS>

F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F:19-131/Domain: cystatin homology <CYI>

F:140-246/Domain: cystatin homology <CY2>

F:276-321/Region: proline-rich

hypothetical prote

enmein precursor

hypothetical prote

protein kinase (EC

zyxin - chicken

protein Y5E21.13 l

genome polypeptide

B3-hordein (clone

hypothetical prote

Krox-20 protein -

WASP, N-WASP, MENA

hypothetical prote

zinc finger transc

DNA strand transfe

insulin-like growt

semaphorin C - mou

Query Match 100.0%; Score 424; DB 1; Length 525;

Best Local Similarity 100.0%; Pred. No. 6.4e-38;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPRPHCRQIGSVYRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPFQSVSESC 60  
|||||  
DB 446 GPRPHCRQIGSVYRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPFQSVSESC 505  
QY 61 GKRKSGFPQVSMFE 75  
|||||  
DB 506 GKRKSGFPQVSMFE 520

RESULT 2  
A60488  
histidine-rich glycoprotein - bovine (fragments)  
N:Alternate names: autorosette inhibition factor  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 19-Mar-1993 #sequence\_revision 23-Mar-1995 #text\_change 07-Jul-1995  
C:Accession: S35687; J02196; A60488  
R:Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.  
PDBS Lett. 328, 285-290, 1993  
A>Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly  
A:Reference number: S35687; MUID:93351678  
A:Accession: S35687

A:Molecule type: protein  
A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>  
A>Note: 355-Gln and 368-Tyr were also found  
R:Halbier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.  
Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
A>Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
A:Reference number: J02196; MUID:94220160  
A:Accession: J02196

A:Molecule type: protein  
A:Residues: 1-23;35-54; 'VK', 57-101, 'R', 'TVGEYREG', 116, 'N', 118, 'R', 120-136;137-145;150-20  
A:Experimental source: plasma  
R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halbier, T.  
Thromb. Res. 60, 385-396, 1990  
A>Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
A:Reference number: A60488; MUID:91196010  
A:Accession: A60488

A:Molecule type: protein  
A:Residues: 1-6, 'X', 8-15 <VES>  
C:Comment: This protein is a single-chained plasma protein which participates in transgl  
C:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: glycoprotein; plasma  
F:2-113/Domain: cystatin homology <CY1>  
F:122-207/Domain: cystatin homology (fragments) <CY2>  
F:7-424,60-71,87-108,165-246,180-203,258-288/Disulfide bonds: #status experimental  
F:74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 58.0%; Score 246; DB 2; Length 445;  
Best Local Similarity 64.3%; Pred. No. 7, 4e-19;  
Matches 45; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

QY 5 FHCROIGSVYRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPFQSVSESCGKFK 64  
|||||  
DB 370 FHCROIGSVYRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPFQSVSESCGKFK 429

QY 65 SGFPQVSMFE 74  
|||||  
DB 430 IKFLHRSKEF 439

RESULT 3  
S18350  
seed storage protein - barley  
C:Species: Hordeum vulgare (barley)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S18350  
R:Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron-Mills, V.  
Plant Mol. Biol. 17, 1217-1231, 1991  
A>Title: Amber codon suppression: the in vivo and in vitro analysis of two C-hordein gen  
A:Reference number: S18350; MUID:92032786  
A:Accession: S18350  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-260 <ENT>  
A:Cross-references: EMBL:X60037; NID:g19000; PIDD:CAA42642.1; PID:g19001  
C:Superfamily: gliadin

Query Match 18.3%; Score 77.5; DB 2; Length 260;  
Best Local Similarity 27.7%; Pred. No. 0.54;  
Matches 26; Conservative 11; Mismatches 28; Indels 29; Gaps 5;

QY 3 RPHCRQ-----IGSVYRLP-----PLRKGEVLPLEAPNPF-----SPFLPH 39  
|||||  
DB 57 QPHTRPQGYRLPPELFPQYQPLPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 116

QY 40 HKHPLKPDNPFQSVSESCGKFKSGFPQVSMFE 73  
|||||  
DB 117 EPIRQPP-QQPFPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 144

RESULT 4  
S09552  
nodulin - soybean  
C:Species: Glycine max (soybean)  
C>Date: 02-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 09-Sep-1997  
C:Accession: S09552  
R:Senupita-Gopalan, C.; Pitas, J.W.; Thompson, D.V.; Hoffman, L.M.  
Mol. Gen. Genet. 203, 410-420, 1986  
A>Title: Expression of host genes during root nodule development in soybeans.  
A:Reference number: S07315  
A:Accession: S09552

A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-360 <SEN>  
A:Cross-references: EMBL:X03979; NID:g18717; PID:g18718

Query Match 18.2%; Score 77; DB 2; Length 360;  
Best Local Similarity 36.2%; Pred. No. 0.88;  
Matches 17; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 10 IGSVYRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPFQ 53  
|||||  
DB 142 LKAVYRLPLRPPPLKTIIFPGTINIVFPRP-----PNIVFPR 181

RESULT 5  
S66312  
G-box binding factor 1A - rape  
C:Species: Brassica napus (rape)  
C>Date: 28-Oct-1996 #sequence\_revision 17-Jul-1998 #text\_change 02-Sep-2000  
C:Accession: S66312; S51306  
R:Waldmuller, S.; Link, G.  
Plant Mol. Biol. 29, 507-517, 1995  
A>Title: Sequence and expression characteristics of three g-box-binding factor cdnas  
A:Reference number: S66310; MUID:96123220  
A:Accession: S66312

A:Molecule type: mRNA  
A:Residues: 1-313 <WAL>  
A:Cross-references: EMBL:X83922; NID:g633149; PIDD:CAA58774.1; PID:g633150  
A:Experimental source: cotyledons: clone pRNGBF  
A>Note: the authors translated 17 amino acids upstream of the initiation codon  
C:Superfamily: fava bean G-box-binding protein; fos/3un DNA-binding domain homology  
C:Keywords: DNA binding; leucine zipper; transcription regulation  
F:232-272/Domain: fos/3un DNA-binding domain homology <FUD>

Query Match 18.0%; Score 76.5; DB 2; Length 313;  
Best Local Similarity 27.6%; Pred. No. 0.85;  
Matches 21; Conservative 8; Mismatches 34; Indels 13; Gaps 2;

QY 2 PRFHCRQIGSV-----YRLPLRKGEVLPLEAPNPFSPPLPHHKHPLKPDNPF 50  
|||||  
DB 41 PSPFPSPVSGSPSPNRYMGAONHMMRY--GTPVRYAMYPGAVYANRPMPPSSAP 98

OY 51 PPOSVSESCPGKFKSG 66  
DB 99 TNETVKEQAPGKSKG 114

## RESULT 6

EWMC  
gamma-gliadin B precursor - wheat  
C:Species: Triticum aestivum (common wheat)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A25632  
R:Rafalski, J.A.  
Gene 43, 221-229, 1986  
A:Title: Structure of wheat gamma-gliadin genes.  
A:Reference number: A25632; MUID:86301876  
A:Accession: A25632  
A:Molecule type: DNA  
A:Residues: 1-291 <RAF>  
A:Cross-references: GB:M13713; NID:g170707; PIDN:AAA34274.1; PID:g170708  
C:Comment: Gliadin, a glutamine and proline rich protein, is a major storage protein for  
C:Superfamily: gliadin  
C:Keywords: seed; storage protein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-291/Product: gamma-gliadin B #status predicted <MAT>

Query Match 17.7%; Score 75; DB 1; Length 291;  
Best Local Similarity 30.2%; Pred. No. 1.1;  
Matches 16; Conservative 7; Mismatches 24; Indels 6; Gaps 1;  
OY 23 EYLPLPEANFPSPFLPHHKLKPD-----NOPPOSVSESCPGKFKSGFPQ 69  
DB 50 QLEPPQQTFFPHQPOQPOQPOQFLQPRQPFPOQPOQPPQPOQPPQ 102

## RESULT 7

C:Chordein precursor - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 11-May-2000  
C:Accession: T06211; A03355  
R:Salnova, O.V.; Mekhedov, S.L.; Zheltnin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.  
Genetika 29, 1070-1079, 1993  
A:Title: Nucleotide sequence of the barley C-hordein gene.  
A:Reference number: Z15337; MUID:93380629  
A:Accession: T06211  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-310 <SAI>  
A:Cross-references: EMBL:S66938; NID:g442523; PIDN:AAB28161.1; PID:g442524  
R:Shewry, P.R.; Autran, J.C.; Nimmo, C.C.; Lew, E.J.L.; Kasarda, D.D.  
Nature 286, 520-522, 1980  
A:Reference number: A93228  
A:Accession: A03355  
A:Molecule type: protein  
A:Residues: 21-48 <SHC>  
C:Superfamily: gliadin  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-310/Product: C-hordein #status experimental <MAT>

Query Match 17.7%; Score 75; DB 2; Length 310;  
Best Local Similarity 26.7%; Pred. No. 1.2;  
Matches 24; Conservative 11; Mismatches 31; Indels 24; Gaps 4;

OY 3 RPFHCRQ-----IGSVYRLP-----PLRKGEVLPLPEANFP-----SPFLPH 39  
DB 57 QPFHPRQQTFFPLPELRSQYQYPLPLQPOQPPQPOQPLPRQPOQPFQPOQPPQ 116  
OY 40 HKHPLKPDNOPPOSVSESCPGKFKSGFPQ 69  
DB 117 QPLPYQP-QQPFNQPOQQLISQOPQOPFPQ 145

RESULT 8  
T35474  
50KD proline rich protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35474  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: Z21579  
A:Accession: T35474  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-456 <COLI>  
A:Cross-references: EMBL:AL034492; PIDN:CAA22501.1; GSPDB:GN00070; SCOEDB:SC6C5.03  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC6C5.03

Query Match 17.7%; Score 75; DB 2; Length 456;  
Best Local Similarity 33.3%; Pred. No. 1.9;  
Matches 17; Conservative 8; Mismatches 14; Indels 12; Gaps 2;

OY 2 PRPFHCRQIGSVYRLPLRKGEVLPLPEANFPSPFLPHHKLKPDNOPPP 52  
DB 199 PRPYG-----RRPERR-----PYPSHVYTRPDPFTRSDFIRPDNRPV 237

## RESULT 9

S68858  
finger protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 01-Dec-2000  
C:Accession: S68858  
R:Perez, M.; Rompato, G.; Corbi, N.; de Gregorio, L.; Draganl, T.A.; Passananti, C.  
FEBS Lett. 387, 117-121, 1996  
A:Title: Zfp60, a mouse zinc-finger gene expressed transiently during in vitro muscle  
A:Reference number: S68858; MUID:96244506  
A:Accession: S68858  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-707 <PER>  
A:Cross-references: GB:U048721; NID:g1216476; PIDN:AAB06876.1; PID:g1216477  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C:Keywords: zinc finger

Query Match 17.5%; Score 74; DB 2; Length 707;  
Best Local Similarity 31.8%; Pred. No. 3.8;  
Matches 27; Conservative 12; Mismatches 30; Indels 16; Gaps 7;

OY 1 GPPFHCROIGSVYRLPL-----RKGEVLPLP-EANF--PSFLP-----HHKHLKPDNQ 49  
DB 197 GEKPYECKDCGKTFRLPQMLSRHKSDBRPFECNICGSHFLPTLLQYHKN-IRHGLK 255  
OY 50 PFPQSVSESCPGKFKSGFPQVSMF 74  
DB 256 PFE---CEECGKSKFS-FNRISTLF 276

## RESULT 10

S18235  
omega secalin precursor (clone pSec1B) - rye  
C:Species: Secale cereale (rye)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 20-Aug-1999  
C:Accession: S18235; S19777  
R:Hull, G.A.; Halford, N.G.; Kreis, M.; Shewry, P.R.  
Plant Mol. Biol. 17, 1111-1115, 1991  
A:Title: Isolation and characterisation of genes encoding rye prolamins containing a  
A:Reference number: S18235; MUID:92032773  
A:Accession: S18235

RESULT 15  
S18236  
omega secalin precursor (clone pSec2B) - rye

C:Species: Secale cereale (rye)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 20-Aug-1999  
C:Accession: S18236; S19778  
R:Hull, G.A.; Halford, N.G.; Kreis, M.; Shewry, P.R.  
Plant Mol. Biol. 17, 1111-1115, 1991  
A:Title: Isolation and characterisation of genes encoding rye prolamins containing a hidden  
A:Reference number: S18235; MUID:92032773  
A:Accession: S18236  
A:Molecule type: DNA  
A:Residues: 1-357 <HUI>  
A:Cross-references: EMBL:X60295; NID:q21203; PIDN:CAA42837.1; PID:q21204  
C:Genetics:  
A:Gene: secl  
C:Superfamily: gliadin  
F:1-19/Domain: signal sequence #status predicted <SRG>  
F:20-357/Product: omega secalin #status predicted <MAT>

	Query Match	Similarity	Score	DB	Length
Best Local	38.6%	Pred. No. 3			
Matches	17	Conservative	6	Mismatches	19
				Indels	2
				Gaps	2

Search completed: May 24, 2002, 16:50:09  
Job time: 314 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 24, 2002, 17:03:28 ; Search time 31.45 seconds  
(without alignments)  
35.703 Million cell updates/sec

Title: US-09-730-379E-5

Sequence: 1 DNGPFGQSVSESCPGKFKSGFPGVSMFFT 29

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	159	100.0	525	1 HRG_HUMAN	P04196 homo sapien
2	83	52.2	396	1 HRG_BOVIN	P33433 bos taurus
3	81	50.9	526	1 HRG_RABIT	O28640 coryctolagus
4	51	32.1	795	1 DEGY_CAEL	O01635 caenorhabdi
5	47	28.6	192	1 SCBP_PENSP	P02635 penaeus sp.
6	47	29.6	485	1 HXKB_YEAST	P04807 saccharomyc
7	47	29.6	1690	1 KFLA_HUMAN	Q12756 homo sapien
8	47	29.6	1695	1 KFLA_MOUSE	P33173 mus musculu
9	46	28.9	251	1 GDBO_WHEAT	P08079 triticum ae
10	46	28.9	291	1 GDBB_WHEAT	P06659 triticum ae
11	46	28.9	439	1 FGL2_HUMAN	Q14314 homo sapien
12	46	28.9	530	1 RA12_HUMAN	O09593 homo sapien
13	46	28.9	777	1 UNCB_CAEL	Q21974 caenorhabdi
14	46	28.9	1482	1 NME2_MOUSE	Q01097 mus musculu
15	46	28.9	1484	1 NME2_RAT	O00960 rattus norv
16	45	28.6	1159	1 NME2_HUMAN	Q13724 homo sapien
17	45	28.6	377	1 N124_SCHRO	O09904 schizosacch
18	45	28.3	377	1 TBX9_CAEL	Q22289 caenorhabdi
19	45	28.3	459	1 CAH9_HUMAN	Q16790 homo sapien
20	45	28.3	829	1 CAD3_HUMAN	P22223 homo sapien
21	44.5	28.0	305	1 HOG1_HORVU	P17990 hordeum vul
22	44.5	28.0	401	1 Y148_METSM	P22344 methanobrev
23	44.5	28.0	642	1 HEMA_INCTA	P07974 influenza c
24	44	27.7	334	1 FLS_EUSGR	O09447 eustoma gra
25	44	27.7	628	1 V7OK_TYMW	P10357 turnip yell
26	44	27.7	628	1 V7OK_TYMW	P28478 turnip yell
27	44	27.7	1027	1 KNG_RAT	P08834 rattus norv
28	44	27.7	1027	1 AFI0_HUMAN	P55197 homo sapien
29	44	27.7	2473	1 TOR2_YEAST	P32600 saccharomyc
30	44	27.4	2477	1 F1NC_RAT	P04937 rattus norv
31	43.5	27.4	316	1 CH18_POPTR	P16061 populus tri
32	43.5	27.4	584	1 COBA_HUMAN	P07357 homo sapien
33	43	27.0	94	1 ITR1_NICSY	Q02214 nicotiana s

34	43	27.0	112	1 CYC_FRIAG	O22642 fritillaria
35	43	27.0	215	1 PCP_BACAM	P46107 bacillus am
36	43	27.0	348	1 FLS_PETHY	O07512 petunia hyb
37	43	27.0	573	1 AMH2_HUMAN	O16671 homo sapien
38	43	27.0	661	1 KNG_MOUSE	O08677 mus musculu
39	43	27.0	1068	1 AFI0_MOUSE	O54826 mus musculu
40	42.5	26.7	422	1 CSD_BORBU	O51111 borrelia bu
41	42.5	26.7	548	1 PPOC_TORBC	O24163 nicotiana t
42	42.5	26.7	2459	1 MAPB_RAT	P15205 rattus norv
43	42.5	26.7	2464	1 MAPB_MOUSE	P14873 mus musculu
44	42.5	26.7	2468	1 MAPB_HUMAN	P46821 homo sapien
45	42	26.4	279	1 ATPG_MYCCE	P47640 mycoplasma

## ALIGNMENTS

```

RESULT 1
ID: HRG_HUMAN          STANDARD:      PRT:      525 AA.
AC  P04196:
DT  20-MAR-1987 (Rel. 04, Created)
DT  20-MAR-1987 (Rel. 04, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Histidine-rich glycoprotein precursor (Histidine-proline rich
DE  glycoprotein) (HRG).
GN  HRG.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCBI_TaxID=9606;
[1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=86216149; PubMed=3011091;
RX  Koide T., Foster D.C., Yoshitake S., Davie E.W.;
RT  "Amino acid sequence of human histidine-rich glycoprotein derived
RT  from the nucleotide sequence of its cDNA.";
RL  Biochemistry 25:2220-2225(1986).
[2]
RP  SEQUENCE FROM N.A.
RA  Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;
RN  Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
[3]
RP  SEQUENCE OF 214-247 FROM N.A.
RX  MEDLINE=94245171; PubMed=8186234;
RA  Remnis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,
RT  Blondin L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;
RT  "Evidence for the absence of Intron H of the histidine-rich
RT  glycoprotein (HRG) gene: genetic mapping and in situ localization of
RT  HRG to chromosome 3q28-q29.";
RL  Genomics 19:195-197(1994).
[4]
RP  SEQUENCE OF 19-27.
RX  TISSUE-Plasma;
RA  Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquati C.,
RA  Sanchez J.-C., James R., Tissot J.-D., Bellqvist B.,
RA  Hochstrasser D.F.;
RT  "Plasma protein map: an update by microsequencing.";
RL  Electrophoresis 13:707-714(1992).
-1- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS
HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
MAY MEDATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
COAGULATION CASCADE.
-1- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDDEM REPEATS OF A 5-
RESIDUE SEQUENCE (GHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.
-1- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M3149; AAA52694.1; -  
DR EMBL: AB005803; BAA21613.1; -  
DR EMBL: Z17218; CAA78925.1; -  
DR PIR: A01287; KGHUGH.  
DR SWISS-2DPAGE: P04196; HUMAN.  
DR MIM: 142640; -  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; cystatin; 1.  
DR SMART: SM00043; CY; 2.  
KW Glycoprotein; Heparin-binding; Repeat; Signal.  
FT CHAIN 1 18  
FT SIGNAL 1 18  
FT DOMAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
FT DOMAIN 137 136 CYSTATIN-LIKE 1.  
FT DOMAIN 276 254 CYSTATIN-LIKE 2.  
FT DOMAIN 350 321 PRO-RICH.  
FT DISULFID 24 504 PRO/HIS-RICH.  
FT DISULFID 78 89 BY SIMILARITY.  
FT DISULFID 105 126 BY SIMILARITY.  
FT DISULFID 203 417 BY SIMILARITY.  
FT CARBOHYD 218 241 BY SIMILARITY.  
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;  
  
Query Match 100.0%; Score 159; DB 1; Length 525;  
Best Local Similarity 100.0%; Pred. No. 1.4e-15;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 DNQPFQSVSESCPGKFKSGFPQVSMFET 29  
Db 492 DNQPFQSVSESCPGKFKSGFPQVSMFET 520  
|||||  
|  
  
RESULT 2  
HRG\_BOVIN STANDARD; PRT; 396 AA.  
ID HRG\_BOVIN  
AC P31433;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)  
DE (HPRG) (Fragment).  
GN HRG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE AND DISULFIDE BONDS.  
RX MEDLINE=93351678; PubMed=8348977;  
RA Soerensen C.B., Krogh-Pedersen H., Petersen T.E.;  
RT "Determination of the disulphide bridge arrangement of bovine  
RT histidine-rich glycoprotein.";  
RT FEBS Lett. 328:285-290(1993).  
RL  
CC -I- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD

CC COAGULATION CASCADE.  
CC -I- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
CC PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-  
CC RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.  
CC -I- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.  
DR PIR: S35687; S35687.  
DR Pfam: PF00031; cystatin; 1.  
KW Glycoprotein; Heparin-binding; Repeat.  
FT DOMAIN 1 102 CYSTATIN-LIKE 1.  
FT DOMAIN 103 169 CYSTATIN-LIKE 2.  
FT DOMAIN 191 238 PRO-RICH.  
FT DOMAIN 243 368 PRO/HIS-RICH.  
FT DISULFID 7 375  
FT DISULFID 56 67  
FT DISULFID 77 92  
FT DISULFID 123 297  
FT DISULFID 137 160  
FT DISULFID 212 242  
FT NON\_CONS 52 53  
FT NON\_CONS 70 70  
FT CARBOHYD 71 72  
FT NON\_CONS 78 79  
FT NON\_CONS 91 91  
FT CARBOHYD 103 104  
FT NON\_CONS 122 122  
FT CARBOHYD 163 164  
FT NON\_CONS 220 220  
FT CARBOHYD 263 264  
FT NON\_CONS 303 304  
FT VARIANT 86 86  
FT VARIANT 309 309  
FT VARIANT 322 322  
SQ SEQUENCE 396 AA; 44470 MW; 128A8223499DE6FC CRC64;  
  
Query Match 52.2%; Score 83; DB 1; Length 396;  
Best Local Similarity 61.5%; Pred. No. 0.00011;  
Matches 16; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 3 QPFQSVSESCPGKFKSGFPQVSMF 28  
Db 365 QAFQSVSESCPGTFNFKFLHSKFF 390  
|||||  
|  
  
RESULT 3  
HRG\_RABBIT STANDARD; PRT; 526 AA.  
ID HRG\_RABBIT  
AC Q28640;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
DE glycoprotein) (HPRG) (Fragment).  
GN HRG.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
RX TISSUE-Serum;  
RX MEDLINE=96229917; PubMed=8639676;  
RA Borza D.-B., Tatum F.M., Morgan W.T.;  
RT "Domain structure and conformation of histidine-proline-rich  
RT glycoprotein.";  
RT Biochemistry 35:1925-1934(1996).  
RL  
CC -I- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD  
CC COAGULATION CASCADE.

```

-1- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
CC PROTEIN HAS MANY INTERNAL REPEATS. 15 TANDEM REPEATS OF A 5-
CC RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH
CC REGION.
CC
CC -1- SIMILARITY: CONTAINS 2 CYPSTATIN-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to license@sib-sib.ch).
CC -----
CC
CC DR EMBL: U32189; AAC48516.1; -.
CC DR InterPro: IPR000010; Cystatin.
CC DR Pfam: Pf00031; Cystatin; 1.
CC DR SMART: SM00043; Cy; 2.
CC KW Glycoprotein; Heparin-binding; Repeat; Signal.
CC FT NON_TER 1 1
CC FT SIGNAL <1 8 POTENTIAL.
CC FT CHAIN 9 526 HISTIDINE-RICH GLYCOPROTEIN.
CC FT DOMAIN 9 126 CYPSTATIN-LIKE 1.
CC FT DOMAIN 127 243 CYPSTATIN-LIKE 2.
CC FT DOMAIN 251 296 PRO-RICH.
CC FT DOMAIN 329 498 PRO/HIS-RICH.
CC FT DISULFID 14 505 BY SIMILARITY.
CC FT DISULFID 68 79 BY SIMILARITY.
CC FT DISULFID 95 116 BY SIMILARITY.
CC FT DISULFID 193 415 BY SIMILARITY.
CC FT DISULFID 207 230 BY SIMILARITY.
CC FT DISULFID 272 302 POTENTIAL.
CC FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SITE 303 304 CLEAVAGE (BY PLASMIN).
CC FT SITE 421 422 CLEAVAGE (BY PLASMIN).
CC SQ SEQUENCE 526 AA; 58877 MW; 810F23D367D93D42 CRC64;

Query Match 50.9% Score 81; DB 1; Length 526;
Best Local Similarity 57.7% Pred. No. 0.00028;
Matches 15; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 3 QPFPQSVSESCPGKFKSGPPOVSMEF 28
| | | | | | | | | | | | | | | |
DB 495 QPFPVASESRCPERENGFEAQLSKFF 520

RESULT 4
DEGY_CAEEL STANDARD; PRT; 795 AA.
ID DEGY_CAEEL
CC 001635;
CC
CC 15-JUL-1998 (Rel. 36, Created)
CC 15-JUL-1998 (Rel. 36, Last sequence update)
CC 16-OCT-2001 (Rel. 40, Last annotation update)
CC Degenerin-like protein ZK770.1 in chromosome 1.
CC ZK770.1.
CC Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
CC Rhabditidae; Pelodermidae; Caenorhabditis.
CC NCBI_taxid=6239;
CC
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-BRISTOL N2;
CC Maggi L., Gattung S., Bartko L.;
CC Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-slb.ch/announce/
CC or send an email to license@1sb-slb.ch).
CC -----
DR EMBL; U97404; AAB93309.1; -.
DR WormPep; ZK770.1; CE15411.
DR InterPro; IPR001873; ASC.
DR Pfam; PF00858; ASC; 1.
DR PRINTS; PR01078; AMINACHANNEL.
DR PROSITE; PS01206; ASC; 1.
KW Hypothetical protein; Ionic channel; Transmembrane; Ion transport;
KW Glycoprotein.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 POTENTIAL.
FT DOMAIN 80 739 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 740 760 POTENTIAL.
FT DOMAIN 761 795 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 795 AA; 90776 MW; 2CAACEF7D41770B54 CRC64;
OY 1 DNOPPOSVSEPCGCKFKSGF 21
Db 539 DEQPPPTLGFSAPIGTFVSSF 559

RESULT
5
SCPB_PENSP ID SCPB_PENSP STANDARD: PRT: 192 AA.
AC P02635; 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sarcoplasmic calcium-binding protein, beta chain (SCP beta chain).
OS Penaeus sp. (Penaeid shrimp).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Unclassified 'Penaeus'.
OX NCBI_Taxid=6688;
RN [1]
RP SEQUENCE.
RX MEDLINE=85030293; PubMed=6490607;
RA Takagi T., Konishi K.;
RT "Amino acid sequence of the beta chain of sarcoplasmic calcium
binding protein (SCP) obtained from shrimp tail muscle.";
RL J. Biochem. 96:59-67(1984).
CC -1- FUNCTION: LIKE PARVALBUMINS, SCP'S SEEM TO BE MORE ABUNDANT IN
FAST CONTRACTING MUSCLES, BUT NO FUNCTIONAL RELATIONSHIP CAN BE
ESTABLISHED FROM THIS DISTRIBUTION.
CC -1- SUBUNIT: SCPS FROM GRAVIFISH, LOBSTER, AND SHRIMP ARE POLYMORPHIC
DIMERS: THREE ISOTYPES (ALPHA-ALPHA, ALPHA-BETA, AND BETA-BETA)
HAVE BEEN IDENTIFIED.
CC -1- MISCELLANEOUS: THE SARCOPLASMIC CALCIUM-BINDING PROTEINS ARE
ABUNDANT IN THE MUSCLE OF ARTHROPODS, MOLLUSCS, ANNELIDS, AND
PROTOCHORDATES.
CC -1- MISCELLANEOUS: THIS PROTEIN HAS THREE FUNCTIONAL CALCIUM-BINDING
SITES: POTENTIAL SITE 4 HAS LOST AFFINITY FOR CALCIUM.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
PRT: A03072; KLSBS.
InterPro; IPR002048; EF-hand.

```

DR pfam: PF00036; efhand: 3.  
 DR SMART: SM00054; Efh: 2.  
 DR PROSITE: PS00018; EF\_HAND: 3.  
 KW Muscle protein; Calcium-binding. Repeat: Acetylation.  
 FT MOD\_RES 1 1 ACETYLATION.  
 FT CA\_BIND 17 28 EF\_HAND 1.  
 FT CA\_BIND 69 80 EF\_HAND 2.  
 FT CA\_BIND 113 124 EF\_HAND 3.  
 FT DOMAIN 140 151 ANCESTRAL CALCIUM SITE 4.  
 SQ SEQUENCE 192 AA: 21967 MW: 838123A824E33445 CRC64;

Query Match 29.6%; Score 47; DB 1; Length 192;  
 Best Local Similarity 37.5%; Pident. No. 8.3;  
 Matches 9; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 5 FPOSSESCHGKFKSGPQYSMFF 28  
 DB 81 FKQAVQKNCCKGKAFNPNAFKVF 104

RESULT 6  
 HKRB\_YEAST  
 ID HKRB\_YEAST STANDARD; PRT; 485 AA.  
 AC P04807;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hexokinase B (EC 2.7.1.1) (Hexokinase PII).  
 GN HKX2 OR HKB OR HEX1 OR YGL253W OR NRBA86.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_Taxid=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86120382; PubMed=3003701;  
 RA Stachelin C., Stachelin J., Swan J., Botstein D., Konigsberg W.;  
 RT "Identification, cloning and sequence determination of the genes  
 RT specifying hexokinase A and B from yeast."  
 RL Nucleic Acids Res. 14:945-963(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86056943; PubMed=3905511;  
 RA Froehlich K.-U., Entian K.-D., Mecke D.;  
 RT "The primary structure of the yeast hexokinase PII gene (HKX2) which  
 RT is responsible for glucose repression."  
 RL Gene 36:105-111(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=97127827; PubMed=8972578;  
 RA Coissac E., Maillet E., Robineau S., Netter P.;  
 RT "Sequence of a 39,411 bp DNA fragment covering the left end of  
 RT chromosome VII of Saccharomyces cerevisiae."  
 RL Yeast 12:1555-1562(1996).  
 RN [4]  
 RP SEQUENCE OF 1-246 FROM N.A.  
 RC STRAIN=M303;  
 RX MEDLINE=93311123; PubMed=8332518;  
 RA Breitwieser W., Price C., Schuster T.;  
 RT "Identification of a gene encoding a novel zinc finger protein in  
 RT Saccharomyces cerevisiae."  
 RL Yeast 9:551-556(1993).  
 RN [5]  
 RP SEQUENCE OF 118-126; 175-184 AND 303-313.  
 RC STRAIN=ATCC 38531 / Y41;  
 RX MEDLINE=95255188; PubMed=7737086;  
 RA Norbeck J., Blomberg A.;  
 RT "Gene linkage of two-dimensional polyacrylamide gel electrophoresis  
 RT resolved proteins from isogene families in Saccharomyces cerevisiae  
 RT by microsequencing of in-gel trypsin generated peptides."  
 RL Electrophoresis 16:149-156(1995).

RN [6]  
 RP PHOSPHORYLATION OF SER-14.  
 RX MEDLINE=94114477; PubMed=8286332;  
 RA Kriegl T.M., Rush J., Vojtek A.B., Clifton D., Fraenkel D.G.;  
 RT "In vivo phosphorylation site of hexokinase 2 in Saccharomyces  
 RT cerevisiae."  
 RL Biochemistry 33:148-152(1994).  
 RN [7]  
 RP PHOSPHORYLATION OF SER-157.  
 RX MEDLINE=97199316; PubMed=9047292;  
 RA Heidrich K., Otto A., Behlke J., Rush J., Wenzel K.W., Kriegl T.;  
 RT "Autophosphorylation-inactivation site of hexokinase 2 in  
 RT Saccharomyces cerevisiae."  
 RL Biochemistry 36:1960-1964(1997).  
 RN [8]  
 RP SEQUENCE OF 1-18, AND PHOSPHORYLATION OF SER-14.  
 RX MEDLINE=98384167; PubMed=9718324;  
 RA Behlke J., Heidrich K., Naumann M., Mueller E.-C., Otto A., Reuter R.,  
 RA Kriegl T.;  
 RT "Hexokinase 2 from Saccharomyces cerevisiae: regulation of oligomeric  
 RT structure by in vivo phosphorylation at serine-14."  
 RL Biochemistry 37:11989-11995(1998).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
 RX MEDLINE=78244654; PubMed=355643;  
 RA Anderson C.M., Stenkamp R.E., Steltz T.A.;  
 RT "Sequencing a protein by X-ray crystallography. II. Refinement of  
 RT yeast hexokinase B co-ordinates and sequence at 2.1-A resolution."  
 RL J. Mol. Biol. 123:15-33(1978).  
 CC -1- FUNCTION: MAIN GLUCOSE PHOSPHORYLATING ENZYME. MAY PLAY A  
 CC REGULATORY ROLE IN BOTH INDUCTION AND REPRESSION OF GENE  
 CC EXPRESSION BY GLUCOSE.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.  
 CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE  
 CC INHIBITION BY ATP.  
 CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING  
 CC ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.  
 CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.  
 CC -1- DATABASE: NAME-worthington enzyme manual;  
 CC WWW="http://www.worthington-biochem.com/manual/HK.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X03483; CAA27203.1; -  
 DR EMBL: M1181; AAA34697.1; -  
 DR EMBL: X94357; CAA64134.1; -  
 DR EMBL: Z72775; CAA96973.1; -  
 DR EMBL: X67787; CAA48003.1; -  
 DR PIR: B23523; KIBYHB.  
 DR PIR: S28555; S28555.  
 DR PDB: 2YHX; 15-JUL-92.  
 DR SWISS-2DPAGE: P04807; YEAST.  
 DR YEPD; 8536; -  
 DR YEPD; 8548; -  
 DR SGD; S0003222; HKX2.  
 DR InterPro: IPR001312; Hexokinase.  
 DR Pfam: PF00349; hexokinase.1.  
 DR PRINTS: PR00475; HEXOKINASE.  
 DR ProDom: PD001109; Hexokinase.1.  
 DR PROSITE: PS00378; HEXOKINASES; 1.  
 DR Transferrase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;  
 KW 3D-structure; Phosphorylation.  
 FT INIT MET 0  
 FT BINDING 110 110 ATP (BY SIMILARITY).  
 FT DOMAIN 151 177 GLUCOSE-BINDING (POTENTIAL).

FT	MOD_RES	14	14	PHOSPHORYLATION.
FT	MOD_RES	157	157	
FT	CONFLICT	28	28	N -> I (IN REF. 1).
FT	CONFLICT	32	32	I -> N (IN REF. 2).
FT	CONFLICT	60	60	G -> V (IN REF. 1).
FT	CONFLICT	196	196	T -> S (IN REF. 1).
FT	CONFLICT	420	421	YN -> ST (IN REF. 2).
FT	CONFLICT	443	444	TS -> PH (IN REF. 2).
FT	CONFLICT	452	452	I -> V (IN REF. 2).
FT	CONFLICT	461	461	A -> P (IN REF. 2).
FT	HELIX	23	33	
FT	HELIX	37	55	
FT	STRAND	65	66	
FT	STRAND	79	86	
FT	STRAND	90	99	
FT	TURN	100	101	
FT	STRAND	102	110	
FT	STRAND	112	112	
FT	TURN	115	118	
FT	TURN	123	123	
FT	HELIX	124	141	
FT	TURN	142	142	
FT	STRAND	150	155	
FT	STRAND	164	164	
FT	TURN	165	166	
FT	STRAND	169	169	
FT	TURN	175	176	
FT	STRAND	187	187	
FT	HELIX	188	199	
FT	STRAND	202	208	
FT	HELIX	210	221	
FT	TURN	223	224	
FT	STRAND	225	231	
FT	STRAND	235	241	
FT	HELIX	243	245	
FT	TURN	251	252	
FT	STRAND	263	266	
FT	TURN	270	275	
FT	HELIX	283	291	
FT	HELIX	299	304	
FT	HELIX	306	322	
FT	TURN	323	324	
FT	TURN	338	339	
FT	TURN	343	343	
FT	HELIX	344	351	
FT	HELIX	358	368	
FT	TURN	369	369	
FT	HELIX	374	395	
FT	HELIX	396	396	
FT	HELIX	397	406	
FT	TURN	407	407	
FT	STRAND	411	416	
FT	TURN	418	421	
FT	TURN	424	425	
FT	HELIX	426	438	
FT	HELIX	445	447	
FT	STRAND	450	454	
FT	TURN	458	460	
FT	HELIX	461	475	
SO	SEQUENCE	485 AA;	53811 MW;	EAA7EC796125324E CRC64;

Query Match 29.6%; Score 47; DB 1; Length 485;

Best Local Similarity 42.9%; Pred. No. 22;

Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

2 NQPPQSVSECPGKFKSGFP 22

139 DEQFQGISERIPLEGFTSFP 159

ID	NAME	STANDARD	PRT	1690 AA
AC	KIF1A_HUMAN			
DT	Q12736;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).			
GN	KIF1A OR ATRV.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=96299637; PubMed=8661001;			
RA	Fu-Long R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;			
RT	"Characterization of a kinesin-related gene ATRV, within the tuberous			
RT	sclerosis locus (TSC1) candidate region on chromosome 9q34.";			
RL	Genomics 33:421-429(1996).			
CC	-1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC			
CC	VESSICLE PRECURSORS (BY SIMILARITY).			
CC	-1- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104			
CC	SUBFAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 PH DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/annouce/">http://www.isb-sib.ch/annouce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X90840; CAA62346.1; "			
DR	HSSP; P17119; 3KAR.			
DR	MIM; 601255; "			
DR	InterPro; IPR000253; FHA_domain.			
DR	InterPro; IPR001849; PH.			
DR	InterPro; IPR001752; Kinesin.			
DR	Pfam; PF00498; FHA; 1.			
DR	Pfam; PF00225; kinesin; 1.			
DR	Pfam; PF00169; PH; 1.			
DR	PRINTS; PR00380; KINESINHEAVY.			
DR	SMART; SM00240; FHA; 1.			
DR	SMART; SM00129; Kisc; 1.			
DR	SMART; SM00233; PH; 1.			
DR	PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.			
DR	PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.			
DR	PROSITE; PS50006; FHA_DOMAIN; 1.			
DR	PROSITE; PS50003; PH_DOMAIN; 1.			
KW	Motor protein; Microtubules; ATP-binding; Coiled coil.			
FT	DOMAIN 1	361		
FT	DOMAIN 366	383		
FT	DOMAIN 429	462		
FT	DOMAIN 516	572		
FT	DOMAIN 622	681		
FT	DOMAIN 801	822		
FT	DOMAIN 1575	1673		
FT	NP_BIND 97	104		
FT	ATP (POTENTIAL).			
QO	SEQUENCE 1690 AA; 191083 MW; DBDDEC784624FB4D CRC64;			

RESULT 7  
KE1A\_HUMAN

Query Match	29.6%	Score 47	DB 1	Length 1690
Best Local Similarity	48.0%	Pred. No. 78		
Matches	12	Conservative	3	Mismatches 8; Indels 2; Gaps 1.
QY	1	DNQPFQSVSSECP--GKFKSGEPQ	23	
		:   :   :   :   :		
db	914	DDQHEKFEQSECPVVGMSRSRSTSQ	938	

```

RESULT      8
KIF1A_MOUSE STANDARD; PRT; 1695 AA.
ID KIF1A_MOUSE
AC P33173; 061770;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Kinesin-like protein KIF1A.
GN KIF1A OR KIF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=95292344; PubMed=7539720;
RA Okada Y., Yamazaki H., Sekine-Aizawa Y., Hirokawa N.;
RT "The neuron-specific kinesin superfamily protein KIF1A is a unique
RT monomeric motor for anterograde axonal transport of synaptic vesicle
RT precursors."
RT Cell 81:769-780(1995).
RN [2]
RP PRELIMINARY SEQUENCE OF 100-247 FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA Hirokawa N.;
RT "Kinesin family in murine central nervous system."
RT J. Cell Biol. 119:1287-1296(1992).
RL -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC
CC VESICLE PRECURSORS.
CC -1- SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
CC TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE
CC TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL, LEVELS ARE
CC CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE
CC PERINUCLEAR AND SYNAPTIC REGIONS.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D29951; BAA06221.1; -.
DR PIR; E44259; E44259.
DR HSSP; P17119; 3KAR.
DR MGD; MGI:108391; Kif1a.
DR InterPro; IPR000253; FHA_domain.
DR InterPro; IPR001849; PH.
DR Pfam; PF00498; FHA; 1.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PRO0380; KINESINHEAV.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
DR PROSITE; PS50006; FHA_DOMAIN1; 1.
DR PROSITE; PS50003; PH_DOMAIN1; 1.
DR Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 361 KINESIN-MOTOR.
FT 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).

```

```

FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1580 1678 PH.
FT NP_BIND 97 104 ATP (BY SIMILARITY).
SO SEQUENCE 1695 AA; 191724 MW; D6EC3B88C9CC6 CRC64;

Query Match 29.6%; Score 47; DB 1; Length 1695;
Best Local Similarity 48.0%; Pred. No. 79;
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

OY 1 DNQPPQSVSESCP--GKFKSGFPQ 23
Db 914 DDQHFKEFQSESCPVYGMRSRGTSG 938

RESULT      9
GDBO_WHEAT STANDARD; PRT; 251 AA.
ID GDBO_WHEAT
AC P08079;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Gamma-gliadin precursor (fragment).
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA Scheets K., Rafalski J.A., Hedgcock C., Soell D.G.;
RT "Heptapeptide repeat structure of a wheat gamma-gliadin."
RT Plant Sci. Lett. 37:221-225(1985).
RL -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M16060; AAA34288.1; -.
DR PIR; PS0094; PS0094.
DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 19
FT CHAIN 20 >251 GAMMA-GLIADIN.
FT NON_TER 251 251
SO SEQUENCE 251 AA; 29054 MW; 5A3F12C36C825EAD CRC64;

Query Match 28.9%; Score 46; DB 1; Length 251;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 3 QPFPQSVSESCPGKFKSGFPQ 23
Db 82 QPFPQSVSESCPGKFKSGFPQ 102

RESULT      10
GDBO_WHEAT STANDARD; PRT; 291 AA.
AC P06659;

```

```

RN [2] SEQUENCE FROM N.A.
RA Yuwarij S., Liu M., Marsden P., Levy G.;
RT "Cloning and characterization of Hfg12: the human counterpart to the
RL mouse gene Fg12.";
RN submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
[3]
RP CHARACTERIZATION.
RX MEDLINE=98309432; PubMed=9647217;
RA Marazzi S., Blum S., Hartmann R., Gundersen D., Schreyer M.,
RA Aggraves S., von Fliehdner V., Pyeala R., Ruegg C.;
RT "Characterization of human fibroblast-like protein
secreted by T lymphocytes.";
RL J. Immunol. 161:138-147(1998).
CC -1- FUNCTION: MAY PLAY A ROLE IN PHYSIOLOGIC LYMPHOCYTE FUNCTIONS AT
CC MCCOLAL SITES.
CC -1- SUBUNIT: HOMOTETRAMER; DISULFIDE-LINKED.
CC -1- TISSUE SPECIFICITY: SECRETED.
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN CYTOTOXIC
CC T-CELLS.
CC -1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).
CC -----
DR DR EMBL; Z36531; CAAB5298.1; -.
DR EMBL; AF104015; AADI0825.1; -.
DR DR EMBL; AF104014; AADI0825.1; JOINED.
DR DR HSSP; P02671; IFZD.
DR DR MIM: 605351; -.
DR DR InterPro: IPR002181; Fibrinogen_C.
DR Dr Pfam: PF00147; fibrinogen_C; 1.
DR DR SMART; SM00186; FBG; 1.
DR DR PROSITE; PS00514; FIBRI_N_AG_C_DOMAIN; 1.
KM KMW T-cell; Glycoprotein; Signal.
FT FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 439 FIBROLEUKIN.
FT DOMAIN 210 435 FIBRINOGEN C-TERMINAL.
FT FT DISULFID 213 242 BY SIMILARITY.
FT FT DISULFID 371 384 BY SIMILARITY.
FT FT CARBOHYD 25 25 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT FT CARBOHYD 179 179 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT FT CARBOHYD 235 235 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT FT CARBOHYD 263 263 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT FT CARBOHYD 336 336 N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ SEQUENCE 439 AA; 50228 MW; DF3A56288EA9E68 CRC64;

Query Match 28.9%; Score 46; DB 1; Length 439;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 9 VSESCPGKFKSGFPQVSM 26
| | | : | | | | | : |
Db 413 VSEAHPGGYGKSSFKEAKM 430

RESULT 12
RA12_HUMAN STANDARD; PRT; 530 AA.
AC Q9Y5P3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retinoic acid-induced protein 2.
GN RA12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99168896; PubMed=10049581;  
 RA Walpole S.M., Hiriyana K.T., Nicolaou A., Bingham E.L., Durham J.,  
 RT Vaidin M., Ross M.T., Yates J.R., Steying P.A., Trump D.;  
 RT Identification and characterization of the human homologue (RA12) of  
 RT a mouse retinoic acid-induced gene in Xp22.2;  
 RL Genomics 55:275-283(1999).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF136587; AAD33688.1; -  
 DR MIM; 300217; -  
 FT DOMAIN 200 253 PRO-RICH.  
 SQ SEQUENCE 530 AA; 57148 MW; 9879EB69DC6188F CRC64;  
 OY 4 PFPQSVSESCPGKFKSGFPQ 23  
 Db 247 PIPPIPVQSVSESKFSSSPK 266  
 Query Match 28.9%; Score 46; DB 1; Length 530;  
 Best Local Similarity 45.0%; Pred. No. 33;  
 Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
 RESULT 13  
 UNCL\_CAEEL STANDARD; PRT; 777 AA.  
 ID UNCL\_CAEEL  
 AC Q21974;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Degenerin UNC-8.  
 GN UNC-8 OR R13A1.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Tavernarakis N., Shreffler W., Wang S.L., Driscoll M.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R., Du Z.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U76402; AAB39734.1; -  
 DR EMBL; UA0798; AAB1473.2; -  
 DR WormDep; R13A1.4; CE26381.  
 DR Intepio; IPR001873; ASC.

DR Pfam; PF00858; ASC; 1.  
 DR PRINTS; PR01078; AMINACHANNEL.  
 DR PROSITE; PS01206; ASC; 1.  
 KW Ionic channel; Transmembrane; Ion transport; Glycoprotein.  
 FT DOMAIN 1 128 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 129 149 POTENTIAL.  
 FT DOMAIN 150 689 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 690 710 POTENTIAL.  
 FT DOMAIN 711 777 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 597 597 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 777 AA; 89127 MW; 9B328C3A24A1E7E6 CRC64;  
 OY 1 DNDPQSVSESCPGKFKSGF 21  
 Db 497 DDEFFPDFTFGSAFTGFISSF 517  
 Query Match 28.9%; Score 46; DB 1; Length 777;  
 Best Local Similarity 42.9%; Pred. No. 49;  
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
 RESULT 14  
 NME2\_MOUSE STANDARD; PRT; 1482 AA.  
 ID NME2\_MOUSE  
 AC Q01097; Q9DCB2;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glutamate [NMDA] receptor subunit epsilon 2 precursor (N-methyl  
 DE D-aspartate receptor subtype 2b) (NR2B) (NMDAR2B).  
 GN GRIN2B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92310564; PubMed=1377365;  
 RA Katsuwada T., Kashiwabuchi N., Mori H., Sakimura K., Kushiya E.,  
 RA Araki K., Meguro H., Masaki H., Kumanishi T., Arakawa M.,  
 RA Mishina M.;  
 RT Molecular diversity of the NMDA receptor channel.";  
 RL Nature 358:36-41(1992).  
 RN [2]  
 RP SEQUENCE OF 1-337 FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE-Brain;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochita H.,  
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boftelli D., Bojunga N., Carinici P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT Functional annotation of a full-length mouse cDNA collection.";



	RESULT	15		
NME2_RAT				
ID	NME2_RAT	STANDARD:	PRT:	1482 AA.
AC	Q00960; 062684;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Glutamate [NMDA] receptor subunit epsilon 2 precursor (N-methyl D-aspartate receptor subtype 2B) (NR2B) (NMDAR2B).			
GN	GRIIN2B.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=92271257; PubMed=1350383;			
RA	Monyer H., Sprengel R., Schoepfer R., Herb A., Higuchi M., Lomeli H., Burnashev N., Sakmann B., Seeburg P.H.;			
RT	"heteromeric NMDA receptors: molecular and functional distinction of subtypes.";			
RL	Science 256:1217-1221(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Sullivan J.M., Traynelis S.F., Chen H., Escobar W., Heinemann S.F., Lipton S.A.;			
RT	"Identification of two cysteine residues that are required for redox modulation of the NMDA subtype of glutamate receptor ";			
RL	submitted (JUN-1994) to the EMBL/Genebank/DBJ databases.			
CC	-1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS WITH HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT SENSITIVITY TO MAGNESIUM, MEDIATED BY GLYCINE.			
CC	-1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/cch or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M91562; AAA4714.1; -			
DR	EMBL; U11419; AAA5054.1; -			
DR	PIR; B43274; B43274.			
DR	HSSP; P19491; 1GR2.			
DR	InterPro; IPRO01320; Ion_glu_receptor.			
DR	InterPro; IPRO01311; SRP_glu_receptor.			
DR	Pfam; PF00060; lfg_chan_1.			
DR	SMART; SM00079; PBpe_1.			
KW	Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;			
RV	Ionic channel; Magnesium; Glycoprotein.			
FT	SIGNAL	1	26	
FT	CHAIN	27	1482	
FT	DOMAIN	27	557	
FT	TRANSMEM	558	577	
FT	DOMAIN	578	599	
FT	TRANSMEM	600	620	
FT	EXTRACELLULAR (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	2 (POTENTIAL).			
FT	GLUTAMATE [NMDA] RECEPTOR SUBUNIT			
FT	EPSILON 2			
FT	1 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	2 (POTENTIAL).			

```
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 892 892 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1224 1224 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1351 1351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1448 1448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1464 1464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1236 1236 E -> K (IN REF. 2).
FT CONFLICT 1430 1430 VT -> SA (IN REF. 2).
SQ SEQUENCE 1482 AA; 166070 MW; AEF8B9DF3C1B0D5D CRC64;
```

```
Query Match 28.9%; Score 46; DB 1; Length 1482;
Best Local Similarity 45.5%; Pred. No. 96;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
```

```
OY 6 POSVSESCPGKFKSGFPQVSMF 27
Db 1319 PRVSILKDKGRFMDGSPYAHMF 1340
```

Search completed: May 24, 2002, 17:03:30  
Job time: 865 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 24, 2002, 16:50:13 ; Search time 64.04 Seconds

(without alignments)  
43.513 Million cell updates/sec

Title: US-09-730-379E-5

Perfect score: 159

Sequence: 1 DNQPFQSVSESCPKRKGFPQVSMFET 29

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR.71:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159	100.0	525	1 KGHUGH	histidine-rich gly
2	83	52.2	445	2 A60488	histidine-rich gly
3	51.5	32.4	1245	2 G88104	protein F40E12.2 l
4	51.5	32.4	1245	2 T31953	hypothetical prote
5	51	32.1	795	2 T34468	hypothetical prote
6	50	31.4	1021	2 T08601	hypothetical prote
7	50	31.4	1023	2 T31669	neural zinc finger
8	50	31.4	1032	2 T14124	neural zinc finger
9	49.5	31.1	518	2 B86299	hypothetical prote
10	49.5	31.1	1174	2 T49868	related to suppres
11	47.5	29.9	527	2 T01019	transport protein
12	47	29.6	192	1 KLSBS	calcium-binding pr
13	47	29.6	486	1 KIRYHB	hexokinase (EC 2.7
14	47	29.6	1018	2 T47300	hypothetical prote
15	47	29.6	1695	2 A56921	kinasin family pro
16	46.5	29.2	514	2 S25009	monosaccharide tra
17	46.5	29.2	2567	2 A49551	filamin, Muller ce
18	46	28.9	251	2 PS0094	gamma-gliadin prec
19	46	28.9	291	1 EFWTG	gamma-gliadin B pr
20	46	28.9	439	2 I37391	fibinogen-like pr
21	46	28.9	737	2 T16737	hypothetical prote
22	46	28.9	770	2 T51024	related to C2H2 zi
23	46	28.9	1482	2 B43274	N-methyl-D-asparta
24	46	28.9	1482	2 I49704	glutamate receptor
25	46	28.9	1484	2 S52086	N-methyl-D-asparta
26	45.5	28.6	384	2 F84717	hypothetical prote
27	45.5	28.6	900	2 T33482	hypothetical prote
28	45.5	28.6	1159	2 S62562	probable nuclear p
29	45.5	28.6	4558	2 C82199	RTX toxin RtxA VC1

30	45	28.3	115	2 T07783	leucoanthocyanidin
31	45	28.3	197	2 AC3343	transporter BME107
32	45	28.3	371	2 T08008	leucoanthocyanidin
33	45	28.3	377	1 S41019	transcription fact
34	45	28.3	459	2 I38013	p54/58N - human
35	45	28.3	829	1 IJHUCP	cadherin 3 precurs
36	44.5	28.0	150	2 A59103	hypothetical prote
37	44.5	28.0	305	2 S08312	gamma-hordein 1 pr
38	44.5	28.0	401	2 S28653	hypothetical prote
39	44.5	28.0	636	2 A12552	transposase a1r801
40	44.5	28.0	1567	2 T03730	antigen containing
41	44	27.7	74	2 T17834	hypothetical prote
42	44	27.7	256	2 S51050	hypothetical prote
43	44	27.7	264	2 C25486	K-kinogen, HMW p
44	44	27.7	360	2 B87286	conserved hypothet
45	44	27.7	469	2 T46930	hypothetical prote

## ALIGNMENTS

RESULT 1  
KGHUGH  
histidine-rich glycoprotein precursor - human  
N:Alternate names: HRG  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #text, change 16-Jun-2000  
C:Accession: A01287; S29669  
R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.  
Biochemistry 25, 2220-2225, 1986  
A>Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu  
A:Reference number: A01287; M01D:86216149  
A:Accession: A01287  
A:Molecule type: mRNA  
A:Residues: 1-525 <KOI>  
A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514  
R:Hennis, B.; Havelaar, A.; Kluft, C.  
submitted to the EMBL Data Library, October 1991  
A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly  
A:Reference number: S29669  
A:Accession: S29669  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 214-247 <HEN>  
A:Cross-references: EMBL:217218; NID:g32453; PIDN:CAA78925.1; PID:g32454  
C:Comment: Although its physiological function is not yet known, HRG does bind heme,  
din, and the lysine-binding site of plasminogen. On the basis of its homology with H  
lood coagulation cascade.  
C:Comment: The amino half of this protein is homologous to the first two cystatin-11s  
could not have inhibitory activity.  
C:Comment: In addition to having a high histidine and proline content, this protein t  
e-rich region.  
C:Genetics:  
A:Gene: GDB:HRG  
A:Cross-references: GDB:120055; OMIM:142640  
A:Map position: 3q27-3q27  
C:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat  
F:1-18/Domain: signal sequence #status predicted <Sig>  
F:19-525/Product: histidine-rich glycoprotein #status predicted <NAT>  
F:19-131/Domain: cystatin homology <CY1>  
F:140-246/Domain: cystatin homology <CY2>  
F:276-321/Region: proline-rich  
F:348-437/Region: histidine-rich  
F:351-497/Region: proline-rich  
F:63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 100.0%; Score 159; DB 1; Length 525;  
Best Local Similarity 100.0%; Pred. No. 3.2e-15;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DNQFPQSVSESCPGKFKSGFPQVSMFE 29  
 DB 492 DNQFPQSVSESCPGKFKSGFPQVSMFE 520

## RESULT 2

A60488  
 histidine-rich glycoprotein - bovine (fragments)  
 N:Alternate names: autorosette inhibition factor  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 19-Mar-1993 #sequence\_revision 23-Mar-1995 #text\_change 07-Jul-1995  
 C:Accession: S35687, J02196, A60488  
 R:Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.  
 FEBS Lett. 328, 285-290, 1993  
 A>Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly  
 A:Reference number: S35687, M0ID:93351678  
 A:Accession: S35687  
 A:Molecule type: protein  
 A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>  
 A:Note: 355-Gln and 368-Tyr were also found  
 R:Halaker, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.  
 Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
 A>Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
 A:Reference number: J02196, M0ID:94220160  
 A:Accession: J02196  
 A:Molecule type: protein  
 A:Residues: 1-23;35-54, 'VK', 57-101, 'R', 'TVGEYREG', 116, 'N', 118, 'R', 120-136;137-145;150-20  
 A:Experimental source: Plasma  
 R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halaker, T.  
 Thromb. Res. 60, 385-396, 1990  
 A>Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
 A:Reference number: A60488; M0ID:91196010  
 A:Accession: A60488  
 A:Molecule type: protein  
 A:Residues: 1-6, 'X', 8-15 <YES>  
 C:Comment: This protein is a single-chained plasma protein which participates in transgl  
 C:Superfamily: histidine-rich glycoprotein; cystatin homology  
 C:Keywords: glycoprotein; plasma  
 F:2-113/Domain: cystatin homology <CY1>  
 F:122-207/Domain: cystatin homology (fragments) <CY2>  
 F:7-424, 60-71, 87-108, 165-346, 180-203, 258-288/Disulfide bonds: #status experimental  
 F:74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 52.2%; Score 83; DB 2; Length 445;  
 Best Local Similarity 61.5%; Pred. No. 0.00031;  
 Matches 16; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 3 QPPQSVSESCPGKFKSGFPQVSMFE 28  
 DB 414 QAFPQSVSESCPGKFKSGFPQVSMFE 439

RESULT 3  
 G88104  
 protein F40E12.2 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C:Accession: G88104  
 R:anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
 A:Reference number: A75000; M0ID:99069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_ele  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: G88104  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1245 <STO>  
 A:Cross-references: GB:chr\_II; PIDN:AB66047.1; PID:g2315574; GSPDB:GN10020; CESP:F40E12.  
 C:Genetics:  
 A:Gene: F40E12.2  
 A:Map position: 2

Query Match 32.4%; Score 51.5; DB 2; Length 1245;  
 Best Local Similarity 52.0%; Pred. No. 35;  
 Matches 13; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

OY 1 DNQFPQSVSESCPGKFKSGFPQV 25  
 DB 1 DHQKF-QSKESLPARRKRAPDKS 24

## RESULT 4

T31953  
 hypothetical protein F40E12.2 - Caenorhabditis elegans (fragment)  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T31953  
 R:Goela, D.  
 submitted to the EMBL data library, July 1997  
 A:Description: The sequence of C. elegans cosmid F40E12.  
 A:Reference number: Z21103  
 A:Accession: T31953  
 A:Status: preliminary; translated from GB/EMBL/DDB  
 A:Molecule type: DNA  
 A:Residues: 1-1245 <GOE>  
 A:Cross-references: EMBL:AF016659; PIDN:AB66047.1; GSPDB:GN00020; CESP:F40E12.2  
 A:Experimental source: strain Bristol N2; clone F40E12  
 C:Genetics:  
 A:Gene: CESP:F40E12.2  
 A:Map position: 2  
 A:Note: Intron positions not resolved (incomplete sequence)

Query Match 32.4%; Score 51.5; DB 2; Length 1245;  
 Best Local Similarity 52.0%; Pred. No. 35;  
 Matches 13; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

OY 1 DNQFPQSVSESCPGKFKSGFPQV 25  
 DB 1 DHQKF-QSKESLPARRKRAPDKS 24

## RESULT 5

T34468  
 hypothetical protein ZK770.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T34468  
 R:Maggi, L.; Gattung, S.; Bartko, L.  
 submitted to the EMBL data library, April 1997  
 A:Description: The sequence of C. elegans cosmid ZK770.  
 A:Reference number: Z21530  
 A:Accession: T34468  
 A:Status: preliminary; translated from GB/EMBL/DDB  
 A:Molecule type: DNA  
 A:Residues: 1-795 <MAG>  
 A:Cross-references: EMBL:U97404; PIDN:AA93309.1; GSPDB:GN00019; CESP:ZK770.1  
 A:Experimental source: strain Bristol N2; clone ZK770  
 C:Genetics:  
 A:Gene: CESP:ZK770.1  
 A:Map position: 1  
 A:Introns: 18/1; 84/2; 125/3; 198/3; 257/3; 337/3; 378/3; 424/3; 477/2; 510/1; 563/3;

Query Match 32.1%; Score 51; DB 2; Length 795;  
 Best Local Similarity 47.6%; Pred. No. 26;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 DNQFPQSVSESCPGKFKSGF 21  
 DB 539 DEQFPDTLGSAPRTGTVSSF 559

RESULT 6  
T08601  
hypothetical protein DG1041 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C:Accession: T08601  
R:Loomis, W.F.; Iranfar, N.  
submitted to the EMBL data library, August 1997  
A:Reference number: Z16451  
A:Accession: T08601  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-1021 <LOO>  
A:Cross-references: EMBL:AF019107; NID:92425053; PID:92425054  
A:Experimental source: strain AX4  
C:Genetics:  
A:Gene: DG1041  
A:Introns: 15/1  
A:Superfamily: slime mold (Dictyostelium discoideum) hypothetical protein DG1041

A:Note: NZF-3  
C:Function:  
A:Description: confers repression on the basal activity of promoters containing the c  
C:Keywords: DNA binding; transcription factor; zinc finger

Best Local Similarity 54.5%; Pred. No. 64;  
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;  
Qy 1 DNOPFQSVSESCPGKFKSGFP 22  
Db 988 NNSVPPSYDESC-CGKGSGFP 1008

RESULT 11  
T01019  
Transport protein homolog YUP8H12R.2 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01019  
R:Theologas, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; Kwart  
Oetner, P.; Davis, R.W.  
Submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.  
A:Reference number: Z14227  
A:Accession: T01019  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-527 <THE>  
A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152576; GSPDB:GN00059; ATSP:YUP8  
A:Experimental source: cultivar Columbia  
A:Genetics:  
A:Gene: ATSP:YUP8H12R.2  
A:Map position: 1

Query Match 29.9%; Score 47.5; DB 2; Length 527;  
Best Local Similarity 36.7%; Pred. No. 55;  
Matches 11; Conservative 2; Mismatches 14; Indels 3; Gaps 1;

Qy 1 DNOPFQSVSE---SCPGRKSGFPQVSMF 27  
Db 92 DYSPHVSIVSEWGLQCGASFYKGLPESSSF 121

RESULT 12  
KLSSBS  
calcium-binding protein beta chain - penaeid shrimp (Penaeus sp.)  
N:Alternate names: sarcoplasmic calcium-binding protein (SCP)  
C:Species: Penaeus sp.  
C>Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 17-Apr-1998  
C:Accession: A03072  
R:Takagi, T.; Konishi, K.  
J. Biochem. 96, 59-67, 1984  
A:Title: Amino acid sequence of the beta chain of sarcoplasmic calcium binding protein  
A:Reference number: A03072; MUID:85030293  
A:Accession: A03072  
A:Molecule type: protein  
A:Residues: 1-192 <RAK>  
C:Comment: This protein has three high-affinity calcium-binding sites; however, by compa  
nally have been a fourth site (residues 140-152) that lost its affinity for calcium due  
C:Comment: The sarcoplasmic calcium-binding proteins are abundant in the muscle of arth  
of either identical or very similar chains. The biological function of these proteins m  
C:Superfamily: calmodulin; calmodulin repeat homology  
C:Keywords: acetylated amino end; calcium binding; duplication; EF hand; muscle  
F:4-36/Domain: calmodulin repeat homology <EF1>  
F:56-88/Domain: calmodulin repeat homology <EF2>  
F:100-132/Domain: calmodulin repeat homology <EF3>  
F:1/Modified site: acetylated amino end (Ala) #status experimental  
F:17,19,21,23,28/Binding site: calcium (Asp, Asp, Phe, Asp) #status predicted  
F:69,71,73,75,80/Binding site: calcium (Asp, Asn, Glu, Glu) #status predicted  
F:113,115,117,119,124/Binding site: calcium (Asp, Asp, Met, Glu) #status predicted

Query Match 29.6%; Score 47; DB 1; Length 192;  
Best Local Similarity 37.5%; Pred. No. 23;  
Matches 9; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 5 FPOSVSESCPGKFKSGFPQVSMF 28

Db 81 FKQAVOKCKGKAPANPNAPKVF 104

RESULT 13  
KIBYHB  
hexokinase (EC 2.7.1.1) B - yeast (Saccharomyces cerevisiae)  
N:Alternate names: HEX1 protein; hexokinase II; hexokinase PII; protein G0556; protei  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Dec-1987 #sequence\_revision 12-Apr-1996 #text\_change 16-Jun-2000  
C:Accession: S61608; B23523; S22430; A23958; S05731; S33656; S64279; A53632; S28555  
R:Colisac, E.; Maillier, E.; Robineau, S.; Netter, P.  
Submitted to the EMBL Data Library, December 1995  
A:Reference number: S61598  
A:Accession: S61608  
A:Molecule type: DNA  
A:Residues: 1-486 <COI>  
A:Cross-references: EMBL:X94357; NID:g1150575; PID:CAA64134.1; PID:g1150586  
R:Stachel, C.; Stachel, J.; Swan, J.; Botstein, D.; Konigsberg, W.  
Nucleic Acids Res. 14, 945-963, 1986  
A:Title: Identification, cloning and sequence determination of the genes specifying h  
A:Reference number: A93649; MUID:86120382  
A:Accession: B23523  
A:Molecule type: DNA  
A:Residues: 1-28, 'I', 30-60, 'V', 62-196, 'S', 198-486 <STA>  
A:Cross-references: EMBL:X03483; NID:g3792; PID:CAA27203.1; PID:g3793  
A:Note: the authors translated the codon GTT for residue 61 as Gly  
A:Accession: S22430  
A:Molecule type: protein  
A:Residues: 2-7, 'I', 30-46; 78-111; 114-119; 166-173; 177-194; 228-236; 250-281; 283-300; 330-  
R:Frohlich, K.; Entian, K.; Mecke, D.  
Gene 36, 105-111, 1985  
A:Title: The primary structure of the yeast hexokinase PII gene (HXK2) which is respo  
A:Reference number: A23958; MUID:86056943  
A:Accession: A23958  
A:Molecule type: DNA  
A:Residues: 1-32, 'N', 34-60, 'V', 62-420, 'S', 423-443, 'P', 446-452, 'V', 454-461, 'P', 463-4  
A:Cross-references: EMBL:M11181  
A:Note: the authors translated the codon GTT for residue 61 as Gly  
R:Schmidt, J.J.; Colowick, S.P.  
Arch. Biochem. Biophys. 158, 471-477, 1973  
A:Title: Identification of a peptide sequence involved in association of subunits of  
A:Reference number: S05731; MUID:74114869  
A:Accession: S05731  
A:Molecule type: protein  
A:Residues: 2-12 <SCH>  
R:Breitwieser, W.; Price, C.; Schuster, T.  
Yeast 9, 551-556, 1993  
A:Title: Identification of a gene encoding a novel zinc finger protein in Saccharomy  
A:Reference number: S33654; MUID:93311123  
A:Accession: S33656  
A:Molecule type: DNA  
A:Residues: 1-247 <BRE>  
A:Cross-references: EMBL:X67787; NID:g3707; PID:CAA48003.1; PID:g3710  
R:Colisac, E.; Maillier, E.; Netter, P.  
Submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64271  
A:Accession: S64279  
A:Molecule type: DNA  
A:Residues: 1-486 <COM>  
A:Cross-references: EMBL:Z72775; NID:g1322930; PID:CAA96973.1; PID:g1322931; GSPDB:G  
A:Experimental source: strain S288C  
R:Kriegel, T.M.; Rush, J.; Vojtek, A.B.; Clifton, D.; Frenkel, D.G.  
Biochemistry 33, 148-152, 1994  
A:Title: In vivo phosphorylation site of hexokinase 2 in Saccharomyces cerevisiae.  
A:Reference number: A53632; MUID:94114477  
A:Accession: A53632  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 13-21 <KRI>  
C:Genetics:  
A:Gene: SGD:HXK2; HEX1; SCT2; HKB; MIPS:YGL253W  
A:Cross-references: SGD:S0003222; MIPS:YGL253W

A:Map position: 7L  
C:Superfamily: hexokinase; hexokinase homology  
C:Keywords: allosteric regulation; ATP; glycolysis; phosphoprotein; phosphotransferase  
F:2-486/Product: hexokinase B #status experimental <Mat>  
F:36-470/Domain: hexokinase homology <HXK>  
F:15/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 29.6%; Score 47; DB 1; Length 486;  
Best Local Similarity 42.9%; Pred. No. 60;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 NQFPQSVSESCPGKFGSGFP 22  
::|||::|||::|||  
Db 140 DEQFPQISEPIPLGTFESFP 160

## RESULT 14

T47300  
hypothetical protein T14K23.130 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47300  
R:NYakatura, G.; Faltmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24458  
A:Accession: T47300  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1018 <NTA>  
A:Cross-references: EMBL:AL132909  
A:Experimental source: cultivar Columbia; BAC clone T14K23  
C:Genetics:  
A:Map position: 3  
A:introns: 46/3; 53/3; 111/1; 146/1; 193/1; 238/1; 321/1; 375/1; 419/1; 534/1; 595/1; 79  
A:Note: T14K23.130

Query Match 29.6%; Score 47; DB 2; Length 1018;  
Best Local Similarity 33.3%; Pred. No. 1.3e+02;  
Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 OPFPQSVSESCPGKFGSGFPQVSM 26  
:|::|||::|||  
Db 388 RPTSEVAELIPDFENGMDIDI 411

## RESULT 15

A56921  
kinesin family protein KTF1A - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 02-Feb-2001  
C:Accession: A56921  
R:Okada, Y.; Yamazaki, H.; Sekine-Aizawa, Y.; Hirokawa, N.  
Cell 81, 769-780, 1995  
A:title: The neuron-specific kinesin superfamily protein KTF1A is a unique monomeric mot  
A:Reference number: A56921; MUID:95292344  
A:Accession: A56921  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1695 <RES>  
A:Cross-references: GB:D29951; NID:G976234; PIDN:BA06221.1; PID:G976235  
C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; pleckstri  
C:Keywords: nucleotide binding; P-loop  
F:6-360/Domain: kinesin motor domain homology <KMOT>  
F:97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 29.6%; Score 47; DB 2; Length 1695;  
Best Local Similarity 48.0%; Pred. No. 2.1e+02;  
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

OY 1 DNQFPQSVSESCPGKFGSGFPQ 23  
|::|||::|||  
Db 914 DDQHFQSESCPGVGMRSQTSQ 938

Search completed: May 24, 2002, 16:50:15  
Job time: 320 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 24, 2002, 16:49:01 ; Search time 49.37 Seconds

(without alignments)  
14.348 Million cell updates/sec

Title: US-09-730-379e-5

Perfect score: 159  
Sequence: 1 DNQPPQSVSESCPGKFKSGFPQVSMFT 29Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfills1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	29.6	486	1	US-07-872-678A-48 Sequence 48, Appl
2	47	29.6	520	4	US-07-861-458C-100 Sequence 100, App
3	46.5	29.2	86	4	US-09-247-155-112 Sequence 112, App
4	46	28.9	1456	1	US-08-026-138E-8 Sequence 8, Appl
5	46	28.9	1482	1	US-08-026-138E-2 Sequence 2, Appl
6	46	28.9	1484	2	US-08-231-193A-56 Sequence 56, Appl
7	46	28.9	1484	2	US-08-486-273A-56 Sequence 56, Appl
8	46	28.9	1484	2	US-08-940-086A-56 Sequence 56, Appl
9	46	28.9	1484	2	US-08-940-035A-56 Sequence 56, Appl
10	45	28.3	256	2	US-08-481-658B-51 Sequence 51, Appl
11	45	28.3	256	2	US-08-477-504A-51 Sequence 51, Appl
12	45	28.3	256	2	US-08-486-756A-51 Sequence 51, Appl
13	45	28.3	256	2	US-08-485-862B-51 Sequence 51, Appl
14	45	28.3	256	3	US-08-487-077A-51 Sequence 51, Appl
15	45	28.3	256	3	US-08-485-863A-51 Sequence 51, Appl
16	45	28.3	256	4	US-08-485-049D-51 Sequence 51, Appl
17	45	28.3	257	4	US-08-787-739-51 Sequence 51, Appl
18	45	28.3	257	4	US-09-178-115-51 Sequence 51, Appl
19	45	28.3	257	4	US-09-177-776-51 Sequence 51, Appl
20	45	28.3	377	3	US-08-787-739-87 Sequence 87, Appl
21	45	28.3	377	4	US-09-178-115-87 Sequence 87, Appl
22	45	28.3	377	4	US-09-177-776-87 Sequence 87, Appl
23	45	28.3	422	3	US-08-335-469-2 Sequence 2, Appl
24	45	28.3	459	2	US-08-481-658B-2 Sequence 2, Appl
25	45	28.3	459	2	US-08-477-504A-2 Sequence 2, Appl
26	45	28.3	459	2	US-08-486-756A-2 Sequence 2, Appl
27	45	28.3	459	2	US-08-485-862B-2 Sequence 2, Appl

28	45	28.3	459	3	US-08-787-739-2	Sequence 2, Appl
29	45	28.3	459	3	US-08-487-077A-2	Sequence 2, Appl
30	45	28.3	459	3	US-08-485-863A-2	Sequence 2, Appl
31	45	28.3	459	4	US-08-485-049D-2	Sequence 2, Appl
32	45	28.3	459	4	US-09-178-115-2	Sequence 2, Appl
33	45	28.3	459	4	US-09-177-776-2	Sequence 2, Appl
34	44	27.7	623	4	US-08-653-740-7	Sequence 7, Appl
35	44	27.7	623	2	US-09-073-594-7	Sequence 7, Appl
36	44	27.7	623	3	US-09-275-925-7	Sequence 7, Appl
37	43.5	27.4	316	4	US-07-791-931-8	Sequence 8, Appl
38	43.5	27.4	584	4	US-08-313-288B-17	Sequence 17, Appl
39	43	27.0	241	2	US-08-379-556A-4	Sequence 4, Appl
40	43	27.0	347	2	US-08-379-556A-2	Sequence 2, Appl
41	43	27.0	2628	4	US-09-413-814-11	Sequence 11, Appl
42	42	26.4	31	6	5212296-1	Patent No. 5212296
43	42	26.4	824	2	US-08-785-310A-7	Sequence 7, Appl
44	42	26.4	824	2	US-08-816-693A-52	Sequence 52, Appl
45	42	26.4	824	3	US-08-885-291-52	Sequence 52, Appl

## ALIGNMENTS

```
RESULT 1
US-07-872-678A-48
Sequence 48, Application US/07872678A
Patent No. 5541060
GENERAL INFORMATION:
APPLICANT: Bell, Graeme, et al.
TITLE OF INVENTION: DETECTION OF EARLY-ONSET
TITLE OF INVENTION: NON-INSULIN-DEPENDENT DIABETES MELLITUS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: Post Office Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELEPHONE/INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-872-678A-48

Query Match 29.6%; Score 47; DB 1; Length 486;
Best local Similarity 42.9%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 2 DNQPPQSVSESCPGKFKSGFP 22
Db 140 DEQFPQGISPIPLGFTFSFP 160
```





```

; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTRO SUBUNITS, DNA
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,273A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ. ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1484 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-273A-56

Query Match      28.9%; Score 46; DB 2; Length 1484;
Best Local Similarity 45.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY      6 POSVSESCPGKFKSGPFQVSMF 27
Db      1319 PRVSILKDKGRFMDGSPYAHMF 1340

RESULT      8
US-08-940-086A-56
; Sequence 56, Application US/08940086A
; Patent No. 6111091
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McCauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/940,086A
; FILING DATE: 29-SEPT-97
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24735-9383C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 450-8400
; TELEFAX: (619) 450-8499
; INFORMATION FOR SEQ. ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1484 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-940-086A-56

Query Match      28.9%; Score 46; DB 3; Length 1484;
Best Local Similarity 45.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY      6 POSVSESCPGKFKSGPFQVSMF 27
Db      1319 PRVSILKDKGRFMDGSPYAHMF 1340

RESULT      9
US-08-940-035A-56
; Sequence 56, Application US/08940035A
; Patent No. 6316611
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McCauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,035A
; FILING DATE: 29-SEPT-97
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383E
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1484 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-940-035A-56

Query Match 28.9%; Score 46; DB 4; Length 1484;  
Best Local Similarity 45.5%; Pred. No. 1.4e+02;  
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 6 POSVSESCPGKFKSGFPVOYSMF 27  
|:|:| |:| |:| |:|  
DB 1319 PRSVSLKDKGRFMDGSPYAHMF 1340

RESULT 10  
US-08-481-658B-51  
Sequence 51, Application US/08481658B  
Patent No. 5955075  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: carbonic anhydrase domain  
US-08-481-658B-51

Query Match 28.3%; Score 45; DB 2; Length 256;  
Best Local Similarity 40.7%; Pred. No. 29;  
Matches 11; Conservative 7; Mismatches 5; Indels 4; Gaps 2;

OY 4 PPOSVSESCPGKFKSGF---POVSMF 27  
|:|:| |:| |:| |:|  
DB 13 PMPR-VSPACAGRFOSPVDIRPOLAAF 38

RESULT 11  
US-08-477-504A-51  
Sequence 51, Application US/08477504A  
Patent No. 5972353  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: carbonic anhydrase domain  
US-08-477-504A-51

Query Match 28.3%; Score 45; DB 2; Length 256;  
Best Local Similarity 40.7%; Pred. No. 29;  
Matches 11; Conservative 7; Mismatches 5; Indels 4; Gaps 2;

OY 4 PPOSVSESCPGKFKSGF---POVSMF 27  
|:|:| |:| |:| |:|  
DB 13 PMPR-VSPACAGRFOSPVDIRPOLAAF 38

RESULT 12  
US-08-486-756A-51  
Sequence 51, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA

US-08 13  
US-08-485-862B-51  
Sequence 51, Application US/08485862B  
Patent No. 5989838  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburo  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863

RESULT 14  
 US-08-487-0774-51  
 Sequence 51, Application US/084870774  
 Patent No. 6059242  
 GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jatomir  
 TITLE OF INVENTION: MN Gene and Protein  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Leona L. Lauder  
 STREET: 6 Matiposa Court  
 CITY: Tiburon  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94920  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487,077A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3H  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-435-2034  
 TELEFAX: 415-435-0727  
 INFORMATION FOR SEQ ID NO: 51:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 256 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 DESCRIPTION: carbonic anhydrase domain  
 US-08-487-0774-51

Query Match	28.3%	Score 45;	DB 3;	Length 256;
Best Local Similarity	40.7%;	Pred. No. 29;		
Matches 11;	Conservative 7;	Mismatches 5;	Indels 4;	Gaps 2



THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:05 ; Search time 136.41 Seconds

(without alignments)  
23.614 Million cell updates/sec

Title: US-09-730-379E-5

Perfect score: 159

Sequence: 1 DNQPFQSVSESCRFKFGFQVSMFET 29

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*

8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*

9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*

10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*

11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*

12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*

13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*

14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*

15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*

16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*

17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*

18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*

19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*

20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*

21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	34.6	592	ABB62801	Drosophila melanog
2	55	34.6	592	ABB67339	Drosophila melanog
3	54	34.0	78	AAO06699	Human polypeptide
4	53	33.3	172	ABB65597	Drosophila melanog
5	52	32.7	298	AAAB42724	Human ORFX ORF2488
6	52	32.7	928	AAAB94034	Human protein sequ
7	52	32.7	928	AAAB67451	Amino acid sequenc
8	49	30.8	73	AAV74149	Human prostate tum
9	49	30.8	84	ABB30767	Peptide #3418 enco
10	49	30.8	84	ABB35944	Peptide #3450 enco
11	49	30.8	84	ABB21353	Protein #3352 enco

12	49	30.8	84	22	AAAB56745	Human brain expres
13	49	30.8	84	22	AAAB69127	Human bone marrow
14	49	30.8	84	22	AAAB16958	Peptide #3392 enco
15	49	30.8	84	22	AAAB29447	Peptide #3484 enco
16	49	30.8	84	22	AAAB04657	Peptide #3339 enco
17	49	30.8	375	22	ABG01713	Novel human diagno
18	48	30.2	402	22	ABG02451	Novel human diagno
19	48	30.2	2570	22	ABG06375	Novel human diagno
20	47	29.6	45	21	AAAB32460	Human secreted pro
21	47	29.6	135	22	ABB67507	Drosophila melanog
22	47	29.6	189	22	AAU29290	Human PRO polypept
23	47	29.6	189	22	AAAB65153	Basic fibroblast g
24	47	29.6	202	22	ABB68029	Drosophila melanog
25	47	29.6	225	21	AAAB32457	Human secreted pro
26	47	29.6	378	22	AAAB51347	S. pastorianus mut
27	47	29.6	486	21	AAAB10456	deg-3 gene product
28	47	29.6	508	14	AAAB42747	Secreted protein 1
29	46.5	29.2	86	20	AAV59681	Human polypeptide
30	46.5	29.2	107	22	AAO02900	Human polypeptide
31	46.5	29.2	163	22	AAO06555	Protoninactarum
32	46	28.9	90	22	AAU54437	Novel human diagno
33	46	28.9	105	22	ABG25499	Human prothrombina
34	46	28.9	439	20	AAAB86235	Drosophila melanog
35	46	28.9	1340	22	ABB67659	NMDA receptor chan
36	46	28.9	1456	15	AAAB49042	Rat NMDA receptor
37	46	28.9	1482	14	AAAB44193	Glutamic acid rece
38	46	28.9	1482	15	AAAB45944	Human N-methyl-D-a
39	46	28.9	1484	15	AAAB66040	Human excitatory a
40	46	28.9	1484	17	AAAB80971	Human NMDA recepto
41	46	28.9	1484	16	AAAB92507	Human N-methyl-D-a
42	46	28.9	1484	20	AAAB87510	Human N-methyl-D-a
43	46	28.9	1484	21	AAAB26236	Human NMDAR2B subu
44	46	28.9	1484	21	AAV56134	Human NMDAR2B subu
45	45.5	28.6	384	21	AAAG20827	Arabidopsis thailia

#### ALIGNMENTS

RESULT 1

ABB62801 standard; Protein; 592 AA.

XX

AC ABB62801;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 15195.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

XX

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

DR N-PSDB; ABL06904.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 15195; 21pp + Sequence Listing; English.  
PS  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157737-AB172072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
SQ Sequence 592 AA;

Query Match 34.6%; Score 55; DB 22; Length 592;  
Best Local Similarity 60.0%; Pred. No. 13;  
Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;  
QY 11 ESCP----GKFKSGFPQVSM 26  
Db 229 escpfilltgkxsglpmvsl 248  
||||| ||||| |||

RESULT 2  
AB157739  
ID AB157739 standard; Protein; 592 AA.  
XX  
XX ABB67339;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 28809.  
XX  
XX Drosophila; developmental biology; cell signaling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX N-PSDB; AB11442.  
XX  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX  
XX Disclosure; SEQ ID NO 28809; 21pp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
XX sequences (AB101840-AB16175) and the encoded proteins  
XX (AB157737-AB172072).  
XX  
XX The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
XX  
SQ Sequence 592 AA;

Query Match 34.6%; Score 55; DB 22; Length 592;  
Best Local Similarity 60.0%; Pred. No. 13;  
Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;  
QY 11 ESCP----GKFKSGFPQVSM 26  
Db 229 escpfilltgkxsglpmvsl 248  
||||| ||||| |||

RESULT 3  
AA006699  
ID AA006699 standard; Protein; 78 AA.  
XX  
XX AA006699;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 20591.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSF-) HYSF INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
XX  
XX N-PSDB; AA186630.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
XX  
XX Claim 20; SEQ ID NO 20591; 1399pp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activating/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
XX Sequence 78 AA;

Query Match 34.0%; Score 54; DB 22; Length 78;

Best Local Similarity 41.7%; Pred. No. 2;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 FPOSVSESCPGKFKSGFPQVSMFF 28  
| : : | | | | | : : |  
Db 11 fqlitlscpsafskfpislyf 34

## RESULT 4

ABB66597  
ID ABB66597 standard; Protein; 172 AA.

XX ABB66597;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 26583.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmacetical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL10700.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX interactions -

XX Disclosure: SEQ ID NO 26583; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC cell in developmental biology and in elucidating cell signalling and

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 172 AA;

Query Match 33.3%; Score 53; DB 22; Length 172;

Best Local Similarity 40.7%; Pred. No. 6.7;

Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

OY 2 NOPFPOSVSESCPGKFKSGFPQVSMFF 28

Db 57 ngefkdavkktcvgkkyeefpamraf 83

## RESULT 5

AAAB42724  
ID AAAB42724 standard; Protein; 298 AA.

AC AAAB42724;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2488 polypeptide sequence SEQ ID NO:4976.

XX Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;

XX vulnery; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;

XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

XX antiviral; antibacterial; antifungal; antineumatic; antihypoid;

XX antinaemic; gene therapy; cancer; proliferative disorder; hypertension;

XX neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

XX cholesterol ester storage; systemic lupus erythematosus; infection;

XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

XX bone damage; cartilage damage; antiinflammatory disease; coagulation;

XX thrombosis; contraceptive.

XX Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76933.

PT Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 4148-4149; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;

CC antihypoid; and antinaemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,

CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 298 AA;

Query Match

32.7%; Score 52; DB 21; Length 298;

Best Local Similarity 42.3%; Pred. No. 17;  
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 OPEPOSVSESCPGKFKSGFPQVSMF 28  
111 1:1 111:1:1  
Db 142 qpfilasileelmgpvssgfservlf 167

# RESULT 6

AAB94034  
ID AAB94034 standard; Protein: 928 AA.

AC AAB94034;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14182.

KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 8; SEQ ID 14182; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

Query Match 32.7%; Score 52; DB 22; Length 928;  
Best Local Similarity 42.3%; Pred. No. 61;  
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 OPEPOSVSESCPGKFKSGFPQVSMF 28  
111 1:1 111:1:1  
Db 344 qpfilasileelmgpvssgfservlf 369

# RESULT 7

AAB67451  
ID AAB67451 standard; Protein: 928 AA.

AC AAB67451;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human chaperone polypeptide.

OS Human; chaperone polypeptide; reproductive disease; prolactin production;

KM infertility; tumour; cancer; Peyronie's disease; eye disorder; glaucoma;

KM conjunctivitis; keratitis; neuromuscular disorder; cystic fibrosis;

KM metabolic disorder; Zellweger syndrome; Addison's disease; iritis;

KM autoimmune disorder; inflammatory disorder; systemic lupus erythematosus;

KM acquired immunodeficiency syndrome; AIDS; asthma; atherosclerosis;

KM cell proliferative disorder; gene therapy.

XX Homo sapiens.

XX WO200109178-A2.

XX 03-AUG-2000; 2000MO-US21313.

XX 03-AUG-1999; 99US-0146908.

XX 22-OCT-1999; 99US-0160924.

XX (INCY-) INCYTE GENOMICS INC.

PI Yue H, Bandman O, Tang YT, Baughn MR, Azimzal Y, Lu DAM;

PI WPI: 2001-159853/16.

DR N-PSDB; AAF54990.

XX New human chaperone proteins and polynucleotides, useful in diagnosing,

XX treating and preventing reproductive, eye, neuromuscular, metabolic,

XX autoimmune or inflammatory disorders -

PS Claim 1; Page 99-100; 102pp; English.

XX The present sequence represents a human chaperone polypeptide. Human

CC chaperone polypeptides and polynucleotides are useful in the diagnosis,

CC treatment and prevention of reproductive (e.g. prolactin production,

CC infertility, endometrial or ovarian tumour, cancer of the breast,

CC prostate or testis, Peyronie's disease), eye (e.g. conjunctivitis,

CC keratitis, iritis, glaucoma), neuromuscular, metabolic (e.g. Zellweger

CC syndrome, Addison's disease, cystic fibrosis), and autoimmune and

CC inflammatory disorders (e.g. systemic lupus erythematosus, acquired

CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis), infectious

CC or viral diseases, and cell proliferative disorders. Chaperone

Query Match 32.7%; Score 52; DB 22; Length 928;  
Best Local Similarity 42.3%; Pred. No. 61;  
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 OPEPOSVSESCPGKFKSGFPQVSMF 28  
111 1:1 111:1:1  
Db 344 qpfilasileelmgpvssgfservlf 369

# RESULT 7

AAB67451  
ID AAB67451 standard; Protein: 928 AA.

AC AAB67451;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human chaperone polypeptide.

OS Human; chaperone polypeptide; reproductive disease; prolactin production;

KM infertility; tumour; cancer; Peyronie's disease; eye disorder; glaucoma;

KM conjunctivitis; keratitis; neuromuscular disorder; cystic fibrosis;

KM metabolic disorder; Zellweger syndrome; Addison's disease; iritis;

KM autoimmune disorder; inflammatory disorder; systemic lupus erythematosus;

KM acquired immunodeficiency syndrome; AIDS; asthma; atherosclerosis;

KM cell proliferative disorder; gene therapy.

XX Homo sapiens.

XX WO200109178-A2.

XX 03-AUG-2000; 2000MO-US21313.

XX 03-AUG-1999; 99US-0146908.

XX 22-OCT-1999; 99US-0160924.

XX (INCY-) INCYTE GENOMICS INC.

PI Yue H, Bandman O, Tang YT, Baughn MR, Azimzal Y, Lu DAM;

PI WPI: 2001-159853/16.

DR N-PSDB; AAF54990.

XX New human chaperone proteins and polynucleotides, useful in diagnosing,

XX treating and preventing reproductive, eye, neuromuscular, metabolic,

XX autoimmune or inflammatory disorders -

PS Claim 1; Page 99-100; 102pp; English.

XX The present sequence represents a human chaperone polypeptide. Human

CC chaperone polypeptides and polynucleotides are useful in the diagnosis,

CC treatment and prevention of reproductive (e.g. prolactin production,

CC infertility, endometrial or ovarian tumour, cancer of the breast,

CC prostate or testis, Peyronie's disease), eye (e.g. conjunctivitis,

CC keratitis, iritis, glaucoma), neuromuscular, metabolic (e.g. Zellweger

CC syndrome, Addison's disease, cystic fibrosis), and autoimmune and

CC inflammatory disorders (e.g. systemic lupus erythematosus, acquired

CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis), infectious

CC or viral diseases, and cell proliferative disorders. Chaperone

OY 3 QPEPQSVSESCPGKFKSGFPQVSMFF 28  
||| : : : ||| : : :  
Db 344 qpfllasllleelmqpvssgfsevrllf 369

## RESULT 8

AA74149  
ID AAY74149 standard; Protein; 73 AA.

AC AAY74149;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor EST fragment derived protein #336.

XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

KW treatment.

XX Homo sapiens.

OS DE19820190-A1.

XX 04-NOV-1999.

PF 28-APR-1998; 98DE-1020190.

PR 28-APR-1998; 98DE-1020190.

XX (META-) METAGEN GRS GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-621386/54.

XX N-PSDB; AA525968.

PT New human nucleic acid sequences from pancreatic tumors, and related

PS proteins -

XX Claim 23; Page 449; 502pp; German.

CC This invention describes novel polypeptides and their encoding nucleic

CC acids derived from human pancreatic tumor tissue which have cytostatic

CC activity. The sequences are also useful in producing pharmaceutical

CC compositions for treatment of pancreatic tumors. AAY73814-Y74252

CC represent protein fragments encoded by the human pancreatic tumor cDNA

CC library derived expressed sequence tag (EST) sequences represented in

CC AA252856-253014.

XX Sequence 73 AA;

XX SQ

Query Match 30.8%; Score 49; DB 20; Length 73;

Best Local Similarity 39.3%; Pred. No. 9.9; 11; Indels 2; Gaps 1;

Matches 11; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

OY 3 QPEPQSV--SESCPGKFKSGFPQVSMFF 28

||| : : : ||| : : :  
Db 38 qpfllasllleelmqpvssgfsevrllf 65

## RESULT 9

ABB30767  
ID ABB30767 standard; Peptide; 84 AA.

AC ABB30767;

DT 01-FEB-2002 (first entry)

DE Peptide #318 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;

KW disease; cancer.

OS Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,

XX useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 13735; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

XX nucleic acid probes for measuring gene expression in a sample derived

XX from human breast and Br 474 cells. The method involves contacting

XX the probes with a collection of detectably labelled nucleic acids

XX bound to each probe of the microarray. The probes are useful for

XX verifying the expression of regions of genomic DNA predicted to

XX encode proteins. They are useful for gene discovery, and for

XX determining predisposition and/or prognosis breast disease. Gene

XX expression analysis is useful for assessing the toxicity of chemical

XX agents on cells. The microarray of this invention presents a far greater

XX diversity of probes for measuring gene expression, with far less bias

XX than expressed sequence tag microarrays. The method is suitable for

XX rapid production of functional information from genomic sequence. The

XX present sequence is a peptide encoded by a single exon nucleic acid

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 84 AA;

XX SQ

Query Match 30.8%; Score 49; DB 22; Length 84;

Best Local Similarity 50.0%; Pred. No. 12; 8; Indels 2; Gaps 1;

Matches 11; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

OY 4 PEPQSV--VSESCPGKFKSGFPQ 23

||| : : : ||| : : :  
Db 63 pvpmsgevgdsapgltslaipq 84

## RESULT 10

ABB35944  
ID ABB35944 standard; Peptide; 84 AA.

AC ABB35944;

DT 04-FEB-2002 (first entry)

DE Peptide #3450 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

KW Homo sapiens.

XX XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver  
 XX  
 PS Claim 27; SEQ ID NO 28579; 639pp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 84 AA;

Query Match 30.8%; Score 49; DB 22; Length 84;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 11; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

QY 4 PFPQS--VSESCPGKFKSGFPQ 23  
 | | | | : | | | | |  
 Db 63 pyvmsgevgdsapglfslafpq 84

RESULT 11  
 ABB21353  
 ID ABB21353 standard; Protein; 84 AA.  
 XX  
 AC ABB21353;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #3352 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-48899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts  
 XX  
 PS Claim 15; SEQ ID NO 23123; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 84 AA;

Query Match 30.8%; Score 49; DB 22; Length 84;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 11; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

QY 4 PFPQS--VSESCPGKFKSGFPQ 23  
 | | | | : | | | | |  
 Db 63 pyvmsgevgdsapglfslafpq 84

RESULT 12  
 AAM56745  
 ID AAM56745 standard; Protein; 84 AA.  
 XX  
 AC AAM56745;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28850.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX



OY 4 PFPQS--VSESCPGKFKSGFPQ 23  
 | | | | : | | | | |  
 Db 63 pvpmsgevgdsapq1fslatfpq 84

## RESULT 15

AAM29447  
 ID AAM29447 standard; Protein; 84 AA.

AC AAM29447;

DT 17-OCT-2001 (first entry)

DE Peptide #3484 encoded by probe for measuring placental gene expression.

KM Probe; microarray; human; placenta; antenatal diagnosis;  
 genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PS Claim 27; SEQ ID No 29716; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

CC see AAI31315-AA157546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

SQ Sequence 84 AA;

OY 4 PFPQS--VSESCPGKFKSGFPQ 23  
 | | | | : | | | | |  
 Db 63 pvpmsgevgdsapq1fslatfpq 84

Query Match 30.8%; Score 49; DB 22; Length 84;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 11; Conservative 1; Mismatches 8; Indels 2; Caps 1;

Search completed: May 24, 2002, 16:48:07  
 Job time: 247 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:02:55 ; Search time 107.9 Seconds  
(without alignments)  
11.223 Million cell updates/sec

Title: US-09-730-379E-6  
Perfect score: 34  
Sequence: 1 ASFRVDR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	91.2	383	10 Q9CAP3	Q9CAP3 Arabidopsis
2	30	88.2	179	16 Q9AZA5	Q9AZA5 caulobacter
3	30	88.2	515	11 Q99PS7	Q99PS7 ratius norv
4	30	88.2	525	11 Q99PS8	Q99PS8 ratius norv
5	30	88.2	525	11 Q99PS6	Q99PS6 mus musculu
6	30	88.2	525	11 Q99PS5	Q99PS5 mus musculu
7	30	88.2	525	11 Q99PS5	Q99PS5 mus musculu
8	30	88.2	589	2 Q9S158	Q9S158 comanonas t
9	30	88.2	799	4 Q60302	Q60302 homo sapien
10	30	88.2	810	4 Q96AA8	Q96AA8 homo sapien
11	30	88.2	1007	13 Q90ZM3	Q90ZM3 gallus galli
12	29	85.3	251	10 Q942B1	Q942B1 oryza sativ
13	29	85.3	308	16 Q9HWZ0	Q9HWZ0 pseudomonas
14	29	85.3	308	5 Q9U964	Q9U964 geodia cydo
15	29	85.3	315	10 Q93YL5	Q93YL5 brassica na
16	29	85.3	383	5 Q18431	Q18431 geodia cydo

17	29	85.3	400	16 Q98FQ7	Q98FQ7 rhizobium 1
18	29	85.3	441	10 Q9MAF5	Q9MAF5 brassica na
19	29	85.3	505	5 Q9U965	Q9U965 geodia cydo
20	29	85.3	699	5 Q27656	Q27656 geodia cydo
21	29	85.3	700	5 Q18433	Q18433 geodia cydo
22	29	85.3	764	2 Q52700	Q52700 rhodobacter
23	29	85.3	1873	10 Q9FGI1	Q9FGI1 arabidopsis
24	28	82.4	186	12 Q9J5B1	Q9J5B1 fowlpox vir
25	28	82.4	290	2 Q93BA8	Q93BA8 klebsiella
26	28	82.4	295	2 Q9AGU2	Q9AGU2 burkholderi
27	28	82.4	295	2 Q934I2	Q934I2 burkholderi
28	28	82.4	295	2 Q934I1	Q934I1 burkholderi
29	28	82.4	295	2 Q934I0	Q934I0 burkholderi
30	28	82.4	295	2 Q932Y0	Q932Y0 burkholderi
31	28	82.4	578	2 Q9RAN0	Q9RAN0 burkholderi
32	28	82.4	578	16 Q85350	Q85350 caulobacter
33	28	82.4	634	5 Q9VQH3	Q9VQH3 dirosophila
34	28	82.4	733	10 Q9FJD1	Q9FJD1 arabidopsis
35	27	79.4	108	5 Q9UIE2	Q9UIE2 leishmania
36	27	79.4	162	4 Q96N82	Q96N82 homo sapien
37	27	79.4	194	4 Q96DE0	Q96DE0 homo sapien
38	27	79.4	226	1 Q24783	Q24783 halobacteri
39	27	79.4	236	2 Q9LAK5	Q9LAK5 corynebacte
40	27	79.4	256	2 Q9L0W3	Q9L0W3 streptomyce
41	27	79.4	283	17 Q9HPD5	Q9HPD5 halobacteri
42	27	79.4	323	2 Q9KYU5	Q9KYU5 streptomyce
43	27	79.4	323	16 Q9I3P5	Q9I3P5 pseudomonas
44	27	79.4	340	2 Q54822	Q54822 streptomyce
45	27	79.4	349	16 P72591	P72591 synechocyst

#### ALIGNMENTS

RESULT	ID	Q9CAP3	PRELIMINARY:	PRT:	383 AA.
AC	Q9CAP3	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DE	01-OCT-2001 (TREMBLrel. 18, Last annotation update)				
DE	HYPOTHETICAL 44.8 KDA PROTEIN.				
GN	TSM46.24.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV.COLUMBIA;				
RX	MEDLINE=21016719; PubMed=11130712;				
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,				
RA	White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,				
RA	Buehler E., Chan A., Chao O., Chen H., Cheuk R.F., Chin C.W.,				
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,				
RA	Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,				
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,				
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,				
RA	Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,				
RA	Landin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,				
RA	Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,				
RA	Miltscher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,				
RA	Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,				
RA	Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,				
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,				
RA	Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,				
RA	Wu D., Xu G., Fraser C.M., Venter J.C., Davis R.W.,				
RT	"Sequence and analysis of chromosome I of the plant Arabidopsis				
RT	thaliana."				
RT	Nature 408:816-820(2000).				
DR	EMBL; AC010704; AAC51665.1; -				
DR	InterPro; IPR001810; F-box.				

DR Pfam: PF00646; F-box; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 383 AA; 44777 MW; D92601C1397D8F8 CRC64;

Query Match  
 Best Local Similarity 91.2%; Score 31; DB 10; Length 383;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 250 SFRVDR 256

RESULT 2  
 ID 09A2A5 PRELIMINARY; PRT; 179 AA.  
 AC 09A2A5; 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE HYPOTHETICAL PROTEIN CC3661.  
 GN CC3661.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 19089 / CB15;  
 RA MEDLINE-21173698; PubMed11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Debey R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
 RA Kolony J.F., Smt J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RA "Complete genome sequence of Caulobacter crescentus."  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RL EMBL: AE006024; AAK25623.1; -.  
 DR TIGR: CC3661; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 179 AA; 19828 MW; AC32E4479CF5B42 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 16; Length 179;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 Db 149 SFRVDR 154

RESULT 3  
 ID 099PS7 PRELIMINARY; PRT; 515 AA.  
 AC 099PS7; 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN 2.  
 GN RNRHG2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;

RT "Molecular diversity of mammalian histidine-rich glycoprotein."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB055896; BAB33093.1; -.  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; CY; 2.  
 SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 11; Length 515;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 171 ASFRVDR 177

RESULT 4  
 ID 099PS8 PRELIMINARY; PRT; 525 AA.  
 AC 099PS8; 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DE 01-OCT-2001 (TREMBLrel. 18, last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN 1.  
 GN RNRHG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB055895; BAB33092.1; -.  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; CY; 2.  
 KW SMART; SM00043; CY; 2.  
 SQ SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 11; Length 525;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 171 ASFRVDR 177

RESULT 5  
 ID 099PS6 PRELIMINARY; PRT; 525 AA.  
 AC 099PS6; 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN.  
 GN MHRG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-LIVER;  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB055897; BAB33094.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF000031; cystatin; 1.  
 DR SMART; SM00043; CY; 2.  
 SQ SEQUENCE 525 AA; 59090 MW; A83E93A439CFB3AC CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 11; Length 525;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 |||||:|  
 DB 171 ASFRVER 177

RESULT 6  
 ID Q9ESB3 PRELIMINARY; PRT; 525 AA.  
 AC Q9ESB3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN.  
 GN HRC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=20307726; PubMed=10849117;  
 RA Huelt M.D., Parish C.R.;  
 RT "Murine histidine-rich glycoprotein: cloning, characterization and  
 RT cellular origin";  
 RL Immunol. Cell Biol. 78:280-287(2000).  
 DR EMBL; AF194028; AAG28416.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 2.  
 SQ SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 11; Length 525;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 |||||:|  
 DB 171 ASFRVER 177

RESULT 7  
 ID Q99PS5 PRELIMINARY; PRT; 525 AA.  
 AC Q99PS5;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).  
 GN MHNG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tsuchida N., Wakabayashi S., Jahnke-Dechent W., Koide T.;  
 RT "Structure of mouse histidine-rich glycoprotein gene";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;  
 RA Strusberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB055898; BAB33095.1; -  
 DR EMBL; BC011168; AAH11168.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 2.  
 SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 11; Length 525;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 |||||:|  
 DB 171 ASFRVER 177

RESULT 8  
 ID Q9S158 PRELIMINARY; PRT; 589 AA.  
 AC Q9S158;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 3-(3-HYDROXYPHENYL)PROPIONATE HYDROXYLASE.  
 GN MHPA.  
 OS Comamonas testosteroni (Pseudomonas testosteroni).  
 OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.  
 OC NCBI\_TaxID=285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TA441;  
 RX MEDLINE=20005595; PubMed=10537203;  
 RA Arai H., Yamamoto T., Ohishi T., Shimizu T., Nakata T., Kudo T.;  
 RT "Genetic organization and characterization of the 3-(3-  
 RT hydroxyphenyl)propionic acid degradation pathway of Comamonas  
 RT testosteroni TA441";  
 RL Microbiology 145:2813-2820(1999).  
 CC -1-COFACITOR: FAD (BY SIMILARITY).  
 DR EMBL; AB024335; BAA82878.1; -  
 DR InterPro: IPR001327; FAD\_PYT\_redox.  
 DR InterPro: IPR000733; flavo\_monooxygenase.  
 DR InterPro: IPR002938; Moxy\_FAD\_binding.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR002114; PTS\_HPr\_ser.  
 DR InterPro: IPR00103; Pyridine\_redox\_2.  
 DR InterPro: IPR003042; Rng\_moxxygenase.  
 DR InterPro: IPR000594; Thif\_family.  
 DR Pfam; PF01494; FAD\_binding\_3; 1.  
 DR Pfam; PF01360; Monooxygenase; 1.  
 DR PRINTS; PR00469; FADPDR.  
 DR PRINTS; PR00469; PNDRTASEII.  
 DR PRINTS; PR00420; RINGMOXGNASE.  
 DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
 KW FAD; Flavoprotein; Oxidoreductase.  
 SQ SEQUENCE 589 AA; 65297 MW; 7CACF905825CB744 CRC64;

Query Match  
 Best Local Similarity 88.2%; Score 30; DB 2; Length 589;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 |||||:|  
 DB 281 AGRVDR 287

RESULT 9  
 060302

ID 060302 PRELIMINARY; PRT; 799 AA.  
 AC 060302;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE KIA00555 PROTEIN.  
 GN KIA00555.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,  
 RA Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 5:31-39(1998).  
 DR EMBL: AB011127; BAA25481.1;  
 DR InterPro: IPR001990; Granin.  
 DR PROSITE: PS00422; GRANIN\_1; UNKNOWN\_1.  
 SQ SEQUENCE 799 AA; 93619 MW; 40CD6861B348F2C4 CRC64;

Query Match 88.2%; Score 30; DB 4; Length 799;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 450 ASFRTRD 456  
 RESULT 10  
 Q96A8 PRELIMINARY; PRT; 810 AA.  
 AC 096A8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SIMILAR TO KIA00555 GENE PRODUCT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC017354; AAH17354.1;  
 SQ SEQUENCE 810 AA; 94933 MW; D81480AD76DA6A10 CRC64;

Query Match 88.2%; Score 30; DB 4; Length 810;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 450 ASFRTRD 456  
 RESULT 11  
 Q90ZN3 PRELIMINARY; PRT; 1007 AA.  
 AC 090ZN3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE LAMININ GAMMA 1 (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Halfert W., Dong S., Balasubramani M., Bier M.E.;  
 RT "Aberrant histogenesis after temporary disruption of the retinal basal  
 RT lamina."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF373841; AAK55397.1;  
 FT NON\_TER 1 1007  
 FT NON\_TER 1 1007  
 SQ SEQUENCE 1007 AA; 110999 MW; CCBFD9659E793IPC CRC64;

Query Match 88.2%; Score 30; DB 13; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 Db 335 SFRVDR 340  
 RESULT 12  
 Q942B1 PRELIMINARY; PRT; 251 AA.  
 AC 0942B1;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE B1060H01.30 PROTEIN.  
 GN B1060H01.30.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriaridaceae; Oryzae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
 RT clone:B1060H01."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003560; BAB68095.1;  
 SQ SEQUENCE 251 AA; 26727 MW; F864DB49AB078164 CRC64;

Query Match 85.3%; Score 29; DB 10; Length 251;  
 Best Local Similarity 85.7%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 Db 137 ASFRFDR 143  
 RESULT 13  
 Q9HWZ0 PRELIMINARY; PRT; 308 AA.  
 AC 09HWZ0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE PROBABLE ATP-BINDING COMPONENT OF ABC TRANSPORTER.  
 GN PA4037.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.

OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gartner R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004820; AAC07424.1;  
 DR InterPro: IPR003593; AAA;  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR InterPro: IPR001687; ATP\_GTP\_A.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR SMART: SM00382; AAA; 1.  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 308 AA; 33980 MW; 8A06420351278B64 CRC64;

Query Match 85.3%; Score 29; DB 16; Length 308;  
 Best Local Similarity 83.3%; Pred. No. 76;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 |||:|  
 Db 21 SFRIDR 26

RESULT 14  
 Q90964  
 ID Q90964 PRELIMINARY; PRT; 313 AA.  
 AC Q90964;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CELL RECOGNITION MOLECULE, SHORT FORM (FRAGMENT).  
 GN CRMS.  
 OS Geodia cydonium (Sponge).  
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;  
 OC Astrophorida; Geodidae; Geodia.  
 OX NCBI\_TaxID=6047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99328902; PubMed=10398802;  
 RA Blumberg B., Diehl-Seifert B., Seack J., Steffen R., Mueller I.M.,  
 RA Mueller W.E.G.;  
 RT "Cloning and expression of new receptors belonging to the  
 RT immunoglobulin superfamily from the marine sponge *Geodia cydonium*.";  
 RL Immunogenetics 49:751-763(1999).  
 DR EMBL: Y18373; CAB52477.1;  
 DR InterPro: IPR003599; IG\_  
 DR InterPro: IPR003600; IG\_like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR Pfam: PF00047; IG\_2.  
 DR SMART: SM00409; IG\_2.  
 DR SMART: SM00410; IG\_like; 1.  
 FT NON\_TER  
 SQ SEQUENCE 313 AA; 33987 MW; 224688776B2D313A CRC64;

Query Match 85.3%; Score 29; DB 5; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVD 6  
 |||||  
 Db 204 ASFRVD 209

RESULT 15  
 Q93YL5  
 ID Q93YL5 PRELIMINARY; PRT; 315 AA.  
 AC Q93YL5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 35.7 KDA PROTEIN (FRAGMENT).  
 GN CERP.  
 OS Brassica napus (rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. WI;  
 RA Brugiere N., Cui Y., Jackman L., Rothstein S.J.;  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. WI;  
 RA Cui Y., Brugiere N., Jackman L., Bi Y.M., Rothstein S.J.;  
 RT "A structural and transcriptional comparative analysis of the *S* locus  
 RT regions in two self-incompatible *Brassica napus* lines.";  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AJ249905; CAC80639.1;  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 315 AA; 35713 MW; 412D8471C9B5180F CRC64;

Query Match 85.3%; Score 29; DB 10; Length 315;  
 Best Local Similarity 83.3%; Pred. No. 78;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 |||:|  
 Db 22 SFRIDR 27

Search completed: May 24, 2002, 17:02:57  
 Job time: 887 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:03:30 ; Search time 31.45 Seconds

(without alignments)  
8.618 Million cell updates/sec

Title: US-09-730-379E-6

Perfect score: 34

Sequence: 1 ASFRVDR 7

Scoring table: BIOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	525	1 HRG_HUMAN	P04196 homo sapien
2	34	100.0	526	1 HRG_RABIT	Q28640 oryctolagus
3	30	88.2	810	1 Y555_HUMAN	Q96a88 homo sapien
4	30	88.2	1607	1 LMGI_MOUSE	P02468 mus musculu
5	30	88.2	1609	1 LMGI_HUMAN	P11047 homo sapien
6	29	85.3	605	1 RTK2_GEOCY	P42159 geodia cydo
7	29	85.3	640	1 INVL_CAPAN	P93761 capsicum an
8	29	85.3	857	1 AD22_MOUSE	Q941v6 mus musculu
9	29	85.3	1756	1 YCF1_PINTH	P41647 pinus thunb
10	28	82.4	290	1 BLO2_KLEOX	P23954 klebsiella
11	28	82.4	299	1 CAH5_MOUSE	P23589 mus musculu
12	28	82.4	416	1 SOXB_RHIME	O87388 rhizobium m
13	27	79.4	123	1 KSI6_TREPA	O83875 treponema p
14	27	79.4	153	1 FLAG_VIBCH	Q9K962 vibrio chol
15	27	79.4	707	1 BMP1_XENLA	P98070 xenopus lae
16	27	79.4	1141	1 HEM2_CAEEL	P55163 caenorhabdi
17	27	79.4	3396	1 POLG_DENIS	P33478 d genome po
18	26	76.5	264	1 SPED_PSEAB	O91577 pseudomonas
19	26	76.5	317	1 Y402_RICPR	O94dc9 rickettsia
20	26	76.5	334	1 DMCI_YEAST	P23453 saccharomyc
21	26	76.5	369	1 GLOX_BACSU	O31184 bacillus su
22	26	76.5	413	1 L756_CAEEL	O11184 caenorhabdi
23	26	76.5	466	1 COO6_SCHPO	O9y729 schizosach
24	26	76.5	580	1 GBM4_USTMA	P87035 ustilago ma
25	26	76.5	602	1 Y06B_CAEEL	P34602 caenorhabdi
26	26	76.5	617	1 C1K5_MOUSE	O61762 mus musculu
27	26	76.5	631	1 CYG2_HUMAN	O73343 homo sapien
28	26	76.5	631	1 GIDA_BUCAP	O51879 buchnera ap
29	26	76.5	631	1 GIDA_VIBCH	O9n944 vibrio chol
30	26	76.5	636	1 INVA_LYCES	P29000 lycopersico
31	26	76.5	705	1 STT3_HUMAN	P46977 homo sapien
32	26	76.5	705	1 STT3_MOUSE	P46978 mus musculu
33	26	76.5	740	1 NMDL_HUMAN	O9n9q1 homo sapien

34	26	76.5	842	1	VGIL_HSVBC	P27599 bovine herp
35	26	76.5	971	1	AMPN_HAEEO	O10737 haemochus
36	26	76.5	1142	1	GIG1_CHICK	O02391 gallus gall
37	26	76.5	1160	1	GIG1_CRICK	O92195 cricetus
38	26	76.5	1171	1	GIG1_RAT	O62638 rattus norv
39	26	76.5	1175	1	GIG1_MOUSE	O61543 mus musculu
40	26	76.5	1179	1	GIG1_HUMAN	O92896 homo sapien
41	26	76.5	1882	1	POL2_TRSVR	P25247 tomato ring
42	26	76.5	3354	1	CADN_HUMAN	O9n251 homo sapien
43	25	73.5	68	1	BGTA_MOMCH	P24076 momordica c
44	25	73.5	106	1	Y12K_FCV06	P28709 feline cali
45	25	73.5	106	1	Y12K_FCV44	P28710 feline cali

## ALIGNMENTS

RESULT 1

ID	HRG_HUMAN	STANDARD	PRT	525 AA.
AC	P04196			
DT	20-MAR-1987 (rel. 04, Created)			
DT	20-MAR-1987 (rel. 04, Last sequence update)			
DT	16-OCT-2001 (rel. 40, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).			
DE	HRG.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA."			
RT	Biochemistry 25:2220-2225(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RT	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,			
RT	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;			
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29."			
RT	Genomics 19:195-197(1994).			
RN	[4]			
RP	SEQUENCE OF 19-27.			
RC	TISSUE=Plasma;			
RA	MEDLINE=93092937; PubMed=1459097;			
RT	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,			
RT	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,			
RT	Hochstrasser D.F.;			
RT	"Plasma protein map: an update by microsequencing."			
RT	Electrophoresis 13:707-714(1992).			
CC	-I- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN, AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN MAY MEDATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD COAGULATION CASCADE.			
CC	-I- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATITIONS OF A 5-RESIDUE SEQUENCE (GHHH, CONSENSUS) FORM A HISTIDINE-RICH REGION.			
CC	-I- SIMILARITY: CONTAINS 2 CYSTEINE-LIKE DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M13149; AAS2694.1; -;  
 DR EMBL: AB005803; BAA21613.1; -;  
 DR EMBL: Z17218; CA78925.1; -;  
 DR PIR: A01287; KGHUGH.  
 DR SWISS-2DPAGE: P04196; HUMAN.  
 DR MIM: 142640; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; Cy; 2.  
 DR Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT STGNL 1 18  
 FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
 FT DOMAIN 19 136 CYSTATIN-LIKE 1.  
 FT DOMAIN 137 254 CYSTATIN-LIKE 2.  
 FT DOMAIN 276 321 PRO-RICH.  
 FT DOMAIN 350 497 PRO/HIS-RICH.  
 FT DISULFID 24 504 BY SIMILARITY.  
 FT DISULFID 78 89 BY SIMILARITY.  
 FT DISULFID 105 126 BY SIMILARITY.  
 FT DISULFID 203 417 BY SIMILARITY.  
 FT DISULFID 218 241 BY SIMILARITY.  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 100.0%; Score 34; DB 1; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ASFRVDR 7  
 11111111  
 Db 173 ASFRVDR 179

RESULT 2  
 HRG\_RABIT STANDARD; PRT; 526 AA.  
 AC 028640;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
 DE glycoprotein) (HPRG) (Fragment).  
 GN HRG.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
 RC TISSUE=serum;  
 RX MEDLINE=96229917; PubMed=8639676;  
 RA Borda D.-B., Tatum F.M., Morgan W.T.;  
 RT "Domain structure and conformation of histidine-proline-rich  
 RT glycoprotein";  
 RL Biochemistry 35:1925-1934 (1996).  
 CC -I- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
 CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
 CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
 CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
 CC HOMOLOG WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
 CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD

CC COAGULATION CASCADE.  
 CC -I- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
 CC PROTEIN HAS MANY INTERNAL REPEATS. 15 TANDEN REPEATS OF A 5-  
 CC RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH  
 CC REGION.  
 CC -I- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.

DR EMBL: U32189; AAC48516.1; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; Cy; 2.  
 DR Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT STGNL 1 1  
 FT CHAIN 9 526  
 FT DOMAIN 9 126 CYSTATIN-LIKE 1.  
 FT DOMAIN 127 243 CYSTATIN-LIKE 2.  
 FT DOMAIN 251 296 PRO-RICH.  
 FT DOMAIN 329 498 PRO/HIS-RICH.  
 FT DISULFID 14 505 BY SIMILARITY.  
 FT DISULFID 68 79 BY SIMILARITY.  
 FT DISULFID 95 116 BY SIMILARITY.  
 FT DISULFID 193 415 BY SIMILARITY.  
 FT DISULFID 207 230 BY SIMILARITY.  
 FT DISULFID 272 302 POTENTIAL.  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 303 304 CLEAVAGE (BY PLASMIN).  
 FT SITE 421 422 CLEAVAGE (BY PLASMIN).  
 SQ SEQUENCE 526 AA; 58877 MW; 810F23D367D93D42 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ASFRVDR 7  
 11111111  
 Db 163 ASFRVDR 169

RESULT 3  
 Y555\_HUMAN STANDARD; PRT; 810 AA.  
 AC 096A8; 060302;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein KIA0555.  
 GN KIA0555.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
 RA Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can



RT code for large proteins in vitro.  
 RL DNA Res. 5:31-39(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be  
 CC produced by alternative splicing.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
 CC frameshift in position 805.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB011127; BAA25481.1; ALT\_FRAME.  
 DR EMBL: BC017354; AAH17354.1;  
 KW Hypothetical protein; Alternative splicing.  
 FT VARSPPLIC 539 559 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 810 AA; 94933 MW; D81480AD76DA6A10 CRC64;  
 OY 1 ASFRVDR 7  
 Db 450 ASFRTRD 456  
 Query Match 88.2%; Score 30; DB 1; Length 810;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 4  
 ID LMGL\_MOUSE STANDARD; PRT: 1607 AA.  
 AC P02468;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).  
 GN LMGL OR LMGL-1 OR LAMB-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88059118; PubMed=3680290;  
 RA Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain has a multidomain structure homologous to the  
 RT B1 chain."  
 RL J. Biol. Chem. 262:17111-17117(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89000737; PubMed=3167041;  
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;  
 RT "Primary structure of the mouse laminin B2 chain and comparison with  
 RT laminin B1."  
 RL Biochemistry 27:5198-5204(1988).  
 RN [3]  
 RP SEQUENCE OF 1-239 FROM N.A.  
 RX MEDLINE=8828071; PubMed=2836421;  
 RA Ogawa K., Burdello P.D., Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain promoter contains unique repeat sequences and  
 RT is active in transient transfection."  
 RL J. Biol. Chem. 263:8384-8389(1988).  
 RN [4]  
 RP SEQUENCE OF 1391-1607 FROM N.A.  
 RX MEDLINE=85051302; PubMed=6209134;

RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
 RT coiled-coil alpha-helix."  
 RL EMBO J. 3:2355-2362(1984).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.  
 RX MEDLINE=96196434; PubMed=8648630;  
 RA Steinfeld J., Mayer U., Timpl R., Huber R.;  
 RT "Crystal structure of three consecutive laminin-type epidermal growth  
 RT factor-like (LE) modules of laminin gamma1 chain harboring the  
 RT nidogen binding site."  
 RL J. Mol. Biol. 257:644-657(1996).  
 RN [6]  
 RP STRUCTURE BY NMR OF 824-881.  
 RX MEDLINE=96196435; PubMed=8648631;  
 RA Baugartner R., Czisch M., Mayer U., Poeschl E., Huber R.,  
 RA Timpl R., Holak T.A.;  
 RT "Structure of the nidogen binding LE module of the laminin gamma1  
 RT chain in solution."  
 RL J. Mol. Biol. 257:658-668(1996).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),  
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),  
 CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
 CC COMPONENT).  
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X05211; CAA28838.1; -  
 DR EMBL: J03484; AAA39405.1; -  
 DR EMBL: J02930; AAA39408.1; -  
 DR EMBL: J03749; AAA39409.1; -  
 DR PIR: A28469; MMSB2.  
 DR PDB: 1KLO; 20-AUG-97.  
 DR PDB: 1TLE; 12-FEB-97.  
 DR MGD: MGT:99914; lamc1.  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001886; LamNT.  
 DR InterPro: IPR000034; Laminin\_B.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR Pfam: PF00052; laminin\_B\_1.  
 DR Pfam: PF00053; laminin\_EGF\_10.  
 DR Pfam: PF00055; laminin\_Nterm; 1.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR ProDom: PD003082; LamNT; 1.  
 DR ProDom: PD003031; Laminin\_B\_1.  
 DR SMART: SM00180; EGF\_Lam; 9.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00281; Lamb; 1.  
 DR SMART: SM00136; LamNT; 1.  
 DR PROSITE: PS00022; EGF\_1; 8.

DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 10.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.  
 FT SIGNAL 1 33  
 FT CHAIN 34 1607  
 FT DOMAIN 34 283 LAMININ GAMMA-1 CHAIN.  
 FT DOMAIN 284 339 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 340 395 LAMININ EGF-LIKE 1.  
 FT DOMAIN 396 442 LAMININ EGF-LIKE 2.  
 FT DOMAIN 443 492 LAMININ EGF-LIKE 3.  
 FT DOMAIN 493 502 LAMININ EGF-LIKE 4.  
 FT DOMAIN 503 687 LAMININ EGF-LIKE 5 (N-TERMINAL).  
 FT DOMAIN 721 721 LAMININ DOMAIN IV.  
 FT DOMAIN 722 770 LAMININ EGF-LIKE 6.  
 FT DOMAIN 771 825 LAMININ EGF-LIKE 7.  
 FT DOMAIN 826 881 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).  
 FT DOMAIN 882 932 LAMININ EGF-LIKE 9.  
 FT DOMAIN 933 980 LAMININ EGF-LIKE 10.  
 FT DOMAIN 981 1028 LAMININ EGF-LIKE 11.  
 FT DOMAIN 1029 1607 LAMININ EGF-LIKE 11.  
 FT DOMAIN 1034 1594 COILED COIL (POTENTIAL).  
 FT DISULFID 340 349 BY SIMILARITY.  
 FT DISULFID 342 365 BY SIMILARITY.  
 FT DISULFID 368 377 BY SIMILARITY.  
 FT DISULFID 380 393 BY SIMILARITY.  
 FT DISULFID 396 408 BY SIMILARITY.  
 FT DISULFID 398 414 BY SIMILARITY.  
 FT DISULFID 416 425 BY SIMILARITY.  
 FT DISULFID 428 440 BY SIMILARITY.  
 FT DISULFID 443 454 BY SIMILARITY.  
 FT DISULFID 445 461 BY SIMILARITY.  
 FT DISULFID 463 472 BY SIMILARITY.  
 FT DISULFID 475 490 BY SIMILARITY.  
 FT DISULFID 722 731 BY SIMILARITY.  
 FT DISULFID 724 738 BY SIMILARITY.  
 FT DISULFID 740 749 BY SIMILARITY.  
 FT DISULFID 752 768 BY SIMILARITY.  
 FT DISULFID 771 779 BY SIMILARITY.  
 FT DISULFID 773 790 BY SIMILARITY.  
 FT DISULFID 793 802 BY SIMILARITY.  
 FT DISULFID 805 823 BY SIMILARITY.  
 FT DISULFID 826 840 BY SIMILARITY.  
 FT DISULFID 828 847 BY SIMILARITY.  
 FT DISULFID 850 859 BY SIMILARITY.  
 FT DISULFID 862 879 BY SIMILARITY.  
 FT DISULFID 882 896 BY SIMILARITY.  
 FT DISULFID 884 903 BY SIMILARITY.  
 FT DISULFID 905 914 BY SIMILARITY.  
 FT DISULFID 917 930 BY SIMILARITY.  
 FT DISULFID 933 945 BY SIMILARITY.  
 FT DISULFID 935 952 BY SIMILARITY.  
 FT DISULFID 954 963 BY SIMILARITY.  
 FT DISULFID 966 978 BY SIMILARITY.  
 FT DISULFID 981 993 BY SIMILARITY.  
 FT DISULFID 983 999 BY SIMILARITY.  
 FT DISULFID 1001 1010 BY SIMILARITY.  
 FT DISULFID 1013 1026 BY SIMILARITY.  
 FT DISULFID 1029 1029 BY SIMILARITY.  
 FT DISULFID 1032 1032 INTERCHAIN (PROBABLE).  
 FT DISULFID 1598 1598 INTERCHAIN (WITH CHAIN BETA-1).  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1020 1020 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1105 1105 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1159 1159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1203 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 216 216 G -> A (IN REF. 3).  
 FT CONFLICT 260 260 E -> D (IN REF. 2).  
 FT CONFLICT 337 337 S -> C (IN REF. 2).  
 FT CONFLICT 447 448 LR -> PS (IN REF. 2).  
 FT CONFLICT 544 544 D -> Y (IN REF. 2).  
 FT CONFLICT 662 662 T -> S (IN REF. 2).  
 FT CONFLICT 886 886 MISSING (IN REF. 2).  
 FT CONFLICT 1158 1158 MISSING (IN REF. 2).  
 FT CONFLICT 1434 1434 V -> A (IN REF. 2).  
 FT CONFLICT 1475 1475 R -> K (IN REF. 4).  
 FT CONFLICT 1576 1576 D -> N (IN REF. 4).  
 SQ SEQUENCE 1607 AA; 177297 MW; 81B7B08E4869F242 CRC64;  
 Oy 2 SFRVDR 7  
 Db 578 SFRVDR 583  
 Query Match 88.2%; Score 30; DB 1; Length 1607;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 5  
 LMGL\_HUMAN STANDARD; PRT; 1609 AA.  
 AC P11047;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).  
 GN LAMC1 OR LAMB2.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP MEDLINE=91093128; PubMed=1985895;  
 RA Kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;  
 RT "Structure of the human laminin B2 chain reveals extensive  
 RT divergence from the laminin B1 chain gene.";  
 RL J. Biol. Chem. 266:221-228(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88198245; PubMed=3360804;  
 RX Pikkarainen T., Kallunki T., Tryggvason K.;  
 RT "Human laminin B2 chain. Comparison of the complete amino acid  
 RT sequence with the B1 chain reveals variability in sequence homology  
 RT between different structural domains.";  
 RL J. Biol. Chem. 263:6751-6758(1988).  
 RN [3]  
 RP SEQUENCE OF 1393-1609 FROM N.A.  
 RX MEDLINE=89169663; PubMed=3234037;  
 RA Fukushima Y., Pikkarainen T., Kallunki T., Eddy R.L., Byers M.G.,  
 RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;  
 RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of  
 RT the gene to chromosome region 14q25->q31.";  
 RL Cytogenet. Cell Genet. 48:137-141(1988).  
 RN [4]  
 RP SEQUENCE OF 1282-1609 FROM N.A.  
 RX MEDLINE=9226129; PubMed=1806043;  
 RA Santos C.L.S., Sabaga J., Brentani R.;  
 RT "Differences in human laminin B2 sequences.";  
 RL DNA Seq. 1:275-277(1991).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE

DIFFERENT POLYPEPTIDE CHAINS ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END. THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN), LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ). SUBCELLULAR LOCATION: Extracellular.

-1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

-1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.

-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-1- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

-----

THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/> OR SEND AN EMAIL TO [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL: M55317.0. AAAS9492.1. JOINED
DR EMBL: M55317.0. AAAS9492.1. JOINED
DR EMBL: M55301.1. AAAS9492.1. JOINED
DR EMBL: M55311.1. AAAS9492.1. JOINED
DR EMBL: M55312.1. AAAS9492.1. JOINED
DR EMBL: M55213.1. AAAS9492.1. JOINED
DR EMBL: M55214.1. AAAS9492.1. JOINED
DR EMBL: M55215.1. AAAS9492.1. JOINED
DR EMBL: M55216.1. AAAS9492.1. JOINED
DR EMBL: M55193.1. AAAS9492.1. JOINED
DR EMBL: M55193.1. AAAS9492.1. JOINED
DR EMBL: M55194.1. AAAS9492.1. JOINED
DR EMBL: M55195.1. AAAS9492.1. JOINED
DR EMBL: M55196.1. AAAS9492.1. JOINED
DR EMBL: M55197.1. AAAS9492.1. JOINED
DR EMBL: M55198.1. AAAS9492.1. JOINED
DR EMBL: M55199.1. AAAS9492.1. JOINED
DR EMBL: M55200.1. AAAS9492.1. JOINED
DR EMBL: M55202.1. AAAS9492.1. JOINED
DR EMBL: M55203.1. AAAS9492.1. JOINED
DR EMBL: M55204.1. AAAS9492.1. JOINED
DR EMBL: M55205.1. AAAS9492.1. JOINED
DR EMBL: M55206.1. AAAS9492.1. JOINED
DR EMBL: M55207.1. AAAS9492.1. JOINED
DR EMBL: M55208.1. AAAS9492.1. JOINED
DR EMBL: M55209.1. AAAS9492.1. JOINED
DR EMBL: M03202.1. AAAS9488.1. -
DR EMBL: M27654.1. AAAS9489.1. -
DR EMBL: X13939.1. CAAS2122.1. -
DR PIR: S13548.1. MHHUB2.
DR HSSP: P02468.1. ILE.
DR MIM: 150280. -
DR InterPro: IPR004089.1. Chemotaxis_transducer
DR InterPro: IPR000561. EGF_1like.
DR InterPro: IPR001886. Lamnt.
DR InterPro: IPR000034.1. Laminin_B.
DR InterPro: IPR002049.1. Laminin_EGF.
DR Pfam: PF000052.1. Laminin_B_1
DR Pfam: PF00055.1. Laminin_EGF_10.
DR Pfam: PF00055.1. Laminin_Nterm_1.
DR PRINTS: PR00011.1. EGF_LAMININ.
DR ProDom: PD002082.1. Lamnt.1.
DR ProDom: PD003031.1. Laminin_B_1.
DR SMART: SM00180.1. EGF_Lam: 10.
DR SMART: SM00281.1. LamB_1.
DR SMART: SK00136.1. Lamnt_1.
DR PROSITE: PS00022.1. EGF_1_8.
DR PROSITE: PS01186.1. EGF_2_2.
DR PROSITE: PS01246.1. LAMININ_TYPE_EGF_11.

```

KM	Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil		
KM	Laminin EGF-like domain; Cell adhesion; Repeat; Signal.		
KT	SIGNAL	1	33
FT	CHAIN	34	1609
FT	CHAIN	34	285
FT	DOMAIN	286	341
FT	DOMAIN	342	397
FT	DOMAIN	398	444
FT	DOMAIN	445	494
FT	DOMAIN	495	504
FT	DOMAIN	505	629
FT	DOMAIN	690	723
FT	DOMAIN	724	772
FT	DOMAIN	773	827
FT	DOMAIN	828	883
FT	DOMAIN	884	934
FT	DOMAIN	935	982
FT	DOMAIN	983	1030
FT	DOMAIN	1030	1609
FT	DOMAIN	1038	1609
FT	DISULFID	286	295
FT	DISULFID	288	305
FT	DISULFID	307	316
FT	DISULFID	319	339
FT	DISULFID	342	351
FT	DISULFID	370	379
FT	DISULFID	382	395
FT	DISULFID	398	410
PT	DISULFID	400	416
FT	DISULFID	418	427
FT	DISULFID	430	442
FT	DISULFID	445	456
FT	DISULFID	447	463
FT	DISULFID	465	474
FT	DISULFID	477	492
FT	DISULFID	724	733
FT	DISULFID	726	740
FT	DISULFID	742	751
FT	DISULFID	754	770
FT	DISULFID	773	781
FT	DISULFID	775	792
FT	DISULFID	795	804
FT	DISULFID	807	825
FT	DISULFID	828	842
FT	DISULFID	830	849
FT	DISULFID	852	861
FT	DISULFID	864	881
FT	DISULFID	884	898
FT	DISULFID	886	905
FT	DISULFID	907	916
FT	DISULFID	919	932
FT	DISULFID	935	947
FT	DISULFID	937	954
FT	DISULFID	956	965
FT	DISULFID	968	980
FT	DISULFID	983	995
FT	DISULFID	985	1001
FT	DISULFID	1003	1012
FT	DISULFID	1015	1028
FT	DISULFID	1031	1031
FT	DISULFID	1034	1034
FT	DISULFID	1600	1600
FT	CARBOHYD	60	60
FT	CARBOHYD	134	134
FT	CARBOHYD	576	576
FT	CARBOHYD	650	650
FT	CARBOHYD	1022	1022
FT	CARBOHYD	1107	1107
FT	CARBOHYD	1161	1161
FT	CARBOHYD	1175	1175
FT	CARBOHYD	1205	1205
FT	CARBOHYD	1223	1223

FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1380 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1395 1395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 212 212 F -> I (IN REF. 2).  
 SQ SEQUENCE 1609 AA; 177606 MW; B098E20FCD97293B CRC64;

Query Match 88.2%; Score 30; DB 1; Length 1609;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SERVDR 7  
 DB 580 SERVDR 585

RESULT 6  
 RTR2\_GEOCY STANDARD; PRT; 605 AA.  
 ID RTR2\_GEOCY  
 AC P42159;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Class II receptor tyrosine kinase (EC 2.7.1.112) (GCTR).  
 GN TK.  
 OS Geodia cydonium (Sponge).  
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;  
 OC Astrophorida; Geodiidae; Geodia.  
 OX NCBI\_Taxid=6047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95004895; PubMed-7920862;  
 RA Schaeche H., Schroeder H.C., Gamulin V., Rinkevich B., Mueller I.M.,  
 RA Mueller W.E.G.;  
 RT "Molecular cloning of a tyrosine kinase gene from the marine sponge  
 Geodia cydonium: a new member belonging to the receptor tyrosine  
 RT kinase class II family.";  
 RT Mol. Membr. Biol. 11:101-107 (1994).  
 RN [2]  
 RP REVISIONS.  
 RA Mueller W.E.G.;  
 RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-  
 CC PROTEIN KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: X72622; CAA51198.1; .  
 DR HSSP: P11362; IFCG.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR InterPro: IPR003600; Iq\_Like.  
 DR InterPro: IPR002011; Receptor\_tyr\_kin\_II.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00047; Iq: 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR SMART: SM00410; Iq\_Like; 1.  
 DR SMART: SM00219; TyrKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.

KW Transferase: Tyrosine-protein kinase; Receptor; Transmembrane;  
 KW ATP-binding: Phosphorylation.  
 FT DOMAIN 1 84  
 FT TRANSMEM 85 105  
 FT DOMAIN 106 605  
 FT DOMAIN 346 605  
 FT NP\_BIND 352 360  
 FT BINDING 393 393  
 FT ACT\_SITE 496 496  
 FT MOD\_RES 527 527  
 SQ SEQUENCE 605 AA; 67772 MW; E4940BDBDFC9066 CRC64;

Query Match 85.3%; Score 29; DB 1; Length 605;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVD 6  
 DB 62 ASFRVD 67

RESULT 7  
 INV1\_CAPAN STANDARD; PRT; 640 AA.  
 ID INV1\_CAPAN  
 AC P93761;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Acid beta-fructofuranosidase AIV-18 (EC 3.2.1.26) (Acid sucrose-6-  
 DE phosphate hydrolase) (Acid Invertase).  
 OS Capsicum annuum (Bell pepper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;  
 OC Asteridae; eunsterids I; Solanales; Solanaceae; Capsicum.  
 OX NCBI\_Taxid=4072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=fruit;  
 RA Choi D., Lee K.-W., Kim S.;  
 RT "Isolation and characterization of acid invertase cDNA clone in Hot  
 RT pepper (Capsicum annuum L.) fruits.";  
 RT J. Plant Biol. 40:298-303 (1997).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
 CC fructofuranoside residues in beta-D-fructofuranosides.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: U87849; AAB48484.1; .  
 DR InterPro: IPR001362; Glyco\_hydro\_32.  
 DR Pfam: PF00251; Glyco\_hydro\_32; 1.  
 DR PROSITE: PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
 KW Hydrolase: Glycosidase; Transmembrane; Glycoprotein.  
 FT TRANSMEM 34 54  
 FT ACT\_SITE 126 126  
 FT CARBOHYD 152 152  
 FT CARBOHYD 219 219  
 FT CARBOHYD 491 491  
 FT CARBOHYD 612 612  
 SQ SEQUENCE 640 AA; 70620 MW; D3C628B7A7E6870B CRC64;

Query Match 85.3%; Score 29; DB 1; Length 640;  
 Best Local Similarity 85.7%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
 DB 471 ASFEVDR 477

RESULT 8  
 AD22\_MOUSE STANDARD: PRT: 857 AA.

AC 09R1V6: 09R1V5: 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE ADAM 22 precursor (A disintegrin and metalloproteinase domain 22).  
 GN ADAM22.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).  
 RC TISSUE=Brain;  
 RA MEDLINE=93365303; PubMed=10433968;  
 RA Sagane K., Yamazaki K., Mizui Y., Tanaka I.;  
 RT "Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23.";  
 RL Gene 236:79-86(1999).  
 CC -! FUNCTION: PROBABLE LIGAND FOR INTEGRIN IN THE BRAIN. THIS IS A NON  
 CC CATALYTIC METALLOPROTEIN-LIKE PROTEIN.  
 CC -! SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -! TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE BRAIN.  
 CC -! PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY)  
 CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -! SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AB009674; BA83382.1; -;  
 DR EMBL: AB009674; BA83383.1; -;  
 DR HSSP: P18619; 1FVL.  
 DR MGD: MGI:1340046; Adam22.  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002870; Pep\_M12B\_Proprep.  
 DR InterPro: IPR001590; Reptolysin.  
 DR InterPro: IPR000130; Zn\_MTPepidase.  
 DR Pfam: PF00200; disintegrin.1.  
 DR Pfam: PF01562; Pep\_M12B\_proprep.1.  
 DR Pfam: PF01421; Reptolysin.1.  
 DR PRINTS: PR00289; Disintegrin.  
 DR ProDom: PD000664; Disintegrin.1.  
 DR SMART: SM00181; EGF.1.  
 DR PROSITE: PS00215; ADAM\_MEPRO.1.  
 DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE: PS0214; DISINTEGRIN\_2; 1.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
 KW Signal; Glycoprotein; Transmembrane; EGF-like domain;  
 KW Alternative splicing  
 FT SIGNAL 1 23 POTENTIAL.  
 FT PROPEP 24 223 BY SIMILARITY.  
 FT CHAIN 24 857 ADAM 22.  
 FT DOMAIN 24 734 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 735 755 POTENTIAL.  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 756 857 METALLOPROTEINASE-LIKE.  
 FT DOMAIN 224 436 DISINTEGRIN-LIKE.  
 FT DOMAIN 442 529 CYS-RICH.  
 FT DOMAIN 533 666 EGF-LIKE.  
 FT DOMAIN 673 710 BY SIMILARITY.  
 FT DISULFID 347 431 POTENTIAL.  
 FT DISULFID 501 514 BY SIMILARITY.  
 FT DISULFID 677 692 BY SIMILARITY.  
 FT DISULFID 686 698 BY SIMILARITY.  
 FT DISULFID 700 709 BY SIMILARITY.  
 FT CARBOHYD 163 163 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 517 517 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 632 632 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 673 673 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT VARSPLIC 766 801 MISSING (IN ISOFORM BETA).  
 SQ SEQUENCE 857 AA; 94740 MW; CB88FB7000208E09 CRC64;

Query Match Score 29: DB 1; Length 857;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVDR 6  
 DB 87 ASFRVDR 92

RESULT 9  
 YCFL\_PINTH STANDARD: PRT: 1756 AA.

AC P41647;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 205.3 kDa protein ycf1 (ORF 1756).  
 GN YCF1.  
 OS Pinus thunbergii (Green pine) (Japanese black pine).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=33350;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95024047; PubMed=7937893;  
 RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,  
 RA Sugita M.;  
 RT "Loss of all ndh genes as determined by sequencing the entire  
 RT chloroplast genome of the black pine Pinus thunbergii.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).  
 CC -! FUNCTION: NOT YET KNOWN.  
 CC -! SIMILARITY: BELONGS TO THE YCF1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: D17510; BA04442.1; -;  
 DR Mendel: 16862; PINTH.ycf1.m16862.  
 KW Chloroplast; Hypochemical protein.  
 SQ SEQUENCE 1756 AA; 205320 MW; 15C9946A8EBE7B56 CRC64;

Query Match Score 29: DB 1; Length 1756;  
 Best Local Similarity 83.3%; Pred. No. 92;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFRVDR 7  
 DB 11111

Db 188 SFRIDR 193

RESULT 10

BLO2\_KLEOX

ID BLO2\_KLEOX STANDARD: PRT: 290 AA.

AC P23954; P71418; (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 40, Last annotation update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

GN Beta-lactamase Oxy-2 precursor (EC 3.5.2.6) (Penicillinase).

OS Klebsiella oxytoca.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Klebsiella.

OX NCBI\_TaxID=571;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SL911;

RX MEDLINE=96431816; PubMed=8834898;

RA Fournier B., Lagrange P.H., Philippson A.;

RT "Beta-lactamase gene promoters of 71 clinical strains of Klebsiella oxytoca.";

RL Antimicrob. Agents Chemother. 40:1988-1994(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=KH11;

RX MEDLINE=99240387; PubMed=10223957;

RA Wu S.W., Dornbusch K., Kronvall G.;

RT "Genetic characterization of resistance to extended-spectrum beta-lactams in Klebsiella oxytoca isolates recovered from patients with septicemia at hospitals in the Stockholm area.";

RL Antimicrob. Agents Chemother. 43:1294-1297(1999).

RN [4]

RP SEQUENCE OF 28-290.

RC STRAIN=D488;

RX MEDLINE=91537461; PubMed=1909282;

RA Reyraud A., Peduzzi J., Barthelemy M., Labia R.;

RT "Cetotaxime-hydrolysing activity of the beta-lactamase of Klebsiella oxytoca D488 could be related to a threonine residue at position 140.";

RL FEMS Microbiol. Lett. 65:185-192(1991).

RN [5]

RP SEQUENCE OF 28-290.

RC STRAIN=HB60;

RX MEDLINE=98122261; PubMed=9462429;

RA Farzanen S., Peduzzi J., Soter L., Reyraud A., Barthelemy M., Labia R.;

RT "Characterization and amino acid sequence of the OXY-2 group beta-lactamase of pl 5.7 isolated from aztreonam-resistant Klebsiella oxytoca strain HB60.";

RL J. Antimicrob. Chemother. 40:789-795(1997).

CC -I- FUNCTION: HYDROLYSES BROAD-SPECTRUM BETA-LACTAM ANTIBIOTICS. CEFPIAZIDIME.

CC -I- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-amino acid.

CC -I- MISCELLANEOUS: STRAIN KH11 OXY-2 IS KNOWN AS OXY-2A.

CC -I- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: Z49084; CAA88908.1; -

DR EMBL: D84548; BAA12695.1; -

DR EMBL: Y17714; CAA42614.1; -

DR HSSP: Q47066; 1BZA.

DR InterPro: IPR001466; Beta\_lactam.

DR InterPro: IPR000871; Beta\_lactam\_A.

DR Pfam: PF00144; beta-lactamase; 1.

DR PRINTS: PR00118; BLACTAMSEA.

DR PROSITE: PS00146; BETA\_LACTAMASE\_A.1.

DR Hydrolyase; Antibiotic resistance; Signal.

KW SIGNAL

FT CHAIN 1 27

FT ACT SITE 28 290

FT BINDING 72 72

FT VARIANT 236 238

FT VARIANT 15 15

FT VARIANT 199 199

FT VARIANT 225 225

FT VARIANT 255 255

SO SEQUENCE 290 AA; 31136 MW; 134C4BC24E999E7 CRC64;

Query Match 82.4%; Score 28; DB 1; Length 290;

Best Local Similarity 71.4%; Pred. No. 24;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7

Db 160 ATFRIDR 166

RESULT 11

CAHS\_MOUSE

ID CAHS\_MOUSE STANDARD: PRT: 299 AA.

AC P23589;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Carbonic anhydrase Va, mitochondrial precursor (EC 4.2.1.1) (Carbonate dehydratase Va) (CA-VA) (CA Y).

DE C5A OR C5A OR C5A OR C5A OR C5A.

GN C5A OR C5A OR C5A OR C5A OR C5A.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BIO-HTT; TISSUE=liver;

RX MEDLINE=90221910; PubMed=2109313;

RA Amor-Gueret M., Levi-Strauss M.;

RT "Nucleotide and derived amino-acid sequence of a cDNA encoding a new mouse carbonic anhydrase.";

RL Nucleic Acids Res. 18:1646-1646(1990).

RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-41.

RC TISSUE=liver;

RX MEDLINE=95024107; PubMed=7937950;

RA Nagao Y., Srinivasan M., Platano J.S., Svendrowski M., Waheed A., Sliw W.S.;

RT "Mitochondrial carbonic anhydrase (isozyme V) in mouse and rat: cDNA cloning, expression, subcellular localization, processing, and tissue distribution.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:10330-10334(1994).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS).

RX MEDLINE=96074624; PubMed=7479916;

RA Borlack-Stodin P.A., Heck R.W., Laipis P.J., Silverman D.N., Christianson D.W.;

RT "Structure determination of murine mitochondrial carbonic anhydrase V at 2.45-A resolution: Implications for catalytic proton transfer and

```

RT Inhibitor design.
RL Proc. Natl. Acad. Sci. U.S.A. 92:10949-10953(1995).
CC -1- FUNCTION: REVERSIBLE HYDRATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: LIVER.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X51971; CAA56233.1; -.
CC DR PIR: S12579; S12579.
CC DR PDB: 1DMX; 03-APR-96.
CC DR PDB: 1DMY; 03-APR-96.
CC DR PDB: 1URT; 11-JAN-97.
CC DR MGD: MGI101946; Carls.
CC DR InterPro: IPR001148; Carb_anhydrase.
CC DR Pfam: PF00194; carb_anhydrase.1.
CC DR ProDom: PD000865; Carb_anhydrase; 1.
CC DR PROSITE: PS00162; Euk_CO2_ANHYDRASE; 1.
CC KM Lyase; Zinc; Mitochondrion; Transist peptide; 3D-structure.
CC FT TRANSIT 1 29 MITOCHONDRION.
CC FT CHAIN 30 299 CARBONIC ANHYDRASE VA.
CC FT METAL 124 124 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 126 126 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 149 149 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT VARIANT 151 153 VHW -> FM (TN REF. 1).
CC FT SEQUENCE 299 AA; 34072 MW; 2698CABA00686151 CRC64;

```

```

Query Match      82.4%; Score 28; DB 1; Length 299;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ASFRVDR 7
Db 288 SSFRLDR 294

RESULT 12
SOXB_RHIME STANDARD; PRT; 416 AA.
AC 087388;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sarcosine oxidase beta subunit (EC 1.5.3.1) (Sarcosine oxidase subunit
DE B).
GN SOXB OR R00085 OR SMC02608.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN 1 SEQUENCE FROM N.A.
RP STRAIN=1021;
RA Powers E.L., Vuyyuru V., Kahn M.L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN 12 SEQUENCE FROM N.A.
RP STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boissard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,

```

```

RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- FUNCTION: CATALYZES THE OXIDATIVE DEMETHYLATION OF SARCOSE TO
CC YIELD GLYCINE, HYDROGEN PEROXIDE AND 5,10-
CC METHYLENETETRAHYDROFOLATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Sarcosine + H(2)O + O(2) = glycine +
CC formaldehyde + H(2)O(2).
CC -1- CORRECTOR: FAD; CONTAINS A COVALENT FMN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: STRONG, TO OTHER SARCOSE OXIDASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF055582; AAC62218.1; -.
CC DR EMBL: AL591782; CAC41472.1; -.
CC DR Oxidoreductase; FAD; Flavoprotein; FMN; Complete proteome.
CC FT NP_BIND 33 FAD (ADP PART) (BY SIMILARITY).
CC FT SEQUENCE 416 AA; 45373 MW; 86FA18AAAD9E324 CRC64;

```

```

Query Match      82.4%; Score 28; DB 1; Length 416;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ASFRVDR 7
Db 391 AAFRLDR 397

RESULT 13
RS16_TREPA STANDARD; PRT; 123 AA.
ID RS16_TREPA
AC 083875;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S16.
GN RPS16 OR TP0905.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN 1 SEQUENCE FROM N.A.
RP STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson C.M., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ullrich T.,
RA McDonald L., Artisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: AE001259; AAC65857.1; -.  
 DR HSSP: P80379; 1EMW.  
 DR TIGR: TP0905; -.  
 DR InterPro: IPR000307; Ribosomal\_S16.  
 DR Pfam: PF00886; Ribosomal\_S16; 1.  
 DR PROSITE: PS00732; RIBOSOMAL\_S16; FALSE\_NEG.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 123 AA; 13611 MW; 6E2E9B09B21E0FB8 CRC64;

Query Match 79.4%; Score 27; DB 1; Length 123;  
 Best Local Similarity 83.3%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 Db 51 SFRVDR 56

RESULT 14  
 FLAG\_VIBCH STANDARD; PRT; 153 AA.  
 AC 09K062; O34224;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein flag.  
 GN FLAG OR VC2141.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwyn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 RN [2]  
 RP SEQUENCE OF 1-70 FROM N.A.  
 RC STRAIN-Classical Ogawa 395 / ATCC 39541 / Serotype O1;  
 RX MEDLINE=98101470; PubMed=9440520;  
 RA Klose K.E., Mekalanos J.J.;  
 RT "Differential regulation of multiple flagellins in Vibrio cholerae.";  
 RL J. Bacteriol. 180:303-316(1998).  
 CC -1- FUNCTION: NOT KNOWN.  
 CC -1- SIMILARITY: TO FLAG IN OTHER VIBRIO SPECIES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE004287; AAF95286.1; -.  
 DR EMBL: AF007122; AAC01558.1; -.  
 DR TIGR: VC2141; -.  
 KW Complete proteome.  
 SQ SEQUENCE 153 AA; 17319 MW; DC95E1700C563C3E CRC64;

Query Match 79.4%; Score 27; DB 1; Length 153;  
 Best Local Similarity 83.3%; Pred. No. 21;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 Db 100 SFRVDR 105

RESULT 15  
 BMP1\_XENLA STANDARD; PRT; 707 AA.  
 ID BMP1\_XENLA  
 AC P98070;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Bone morphogenetic protein 1 precursor (EC 3.4.24.-) (BMP-1).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=6335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=94085787; PubMed=8262384;  
 RA Maeno M., Xue Y., Wood T.I., Ong R.C., Kung H.F.;  
 RT "Cloning and expression of cDNA encoding Xenopus laevis bone  
 RT morphogenetic protein-1 during early embryonic development.";  
 RL Gene 134:257-261(1993).  
 CC -1- FUNCTION: INVOLVED IN PATTERN FORMATION IN GASTRULA AND LATER  
 CC DIFFERENTIATION OF DEVELOPING ORGANS.  
 CC -1- DEVELOPMENTAL STAGE: BLASTULA, EARLY GASTRULA AND HATCHED  
 CC TADPOLES; LITTLE OR NO EXPRESSION IN MORULA AND LATE GASTRULA.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC  
 CC METALLOPROTEASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 CUB DOMAINS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L12249; AAL6313.1; -.  
 DR HSSP: P00736; IAPQ.  
 DR MEROPS: M12.005; -.  
 DR InterPro: IPR001506; Astacin.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001310; Zn\_MTPetase.  
 DR Pfam: PF01400; Astacin; 1.  
 DR Pfam: PF00431; CUB; 3.  
 DR Pfam: PF00008; EGF-1.  
 DR PRINTS: PR00480; ASTACIN.  
 DR SMART: SM00042; CUB; 3.  
 DR SMART: SM00179; EGF\_Ca; 1.  
 DR SMART: SM00235; ZMGC; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE: PS01180; CUB; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_Ca; 1.  
 KW Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrolase; Protease;  
 KW Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;  
 KW Glycoprotein.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT PROPEP ? 83 POTENTIAL.  
 FT CHAIN 84 707 BONE MORPHOGENETIC PROTEIN 1.  
 FT DOMAIN 84 284 METALLOPROTEASE.



```

FT DOMAIN 285 397 CUB 1.
FT DOMAIN 398 509 CUB 2.
FT DOMAIN 510 551 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 554 666 CUB 3.
FT METAL 176 176 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 177 177 BY SIMILARITY.
FT METAL 180 180 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 186 186 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 514 526 BY SIMILARITY.
FT DISULFID 522 535 BY SIMILARITY.
FT DISULFID 537 550 BY SIMILARITY.
FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 707 AA; 80673 MM; 1B6980D716DC9B8D CRC64;

```

```

Query Match 79.4%; Score 27; DB 1; Length 707;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ASFRVDR 7
   1:|::||
DB 38 ANFKIDR 44

```

```

Search completed: May 24, 2002, 17:03:31
Job time: 866 sec

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:50:15 ; Search time 64.04 Seconds  
(without alignments)  
10.503 Million cell updates/sec

Title: US-09-730-379E-6

Perfect score: 34

Sequence: 1 ASFRVDR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	525	1 KGHUGH	histidine-rich gly
2	31	91.2	383	2 B6806	hypothetical prote
3	30	88.2	179	3 C87703	conserved hypotnet
4	30	88.2	588	2 AD2445	hypothetical prote
5	30	88.2	799	2 T00331	hypothetical prote
6	30	88.2	1607	1 MMMSB2	laminin gamma-1 ch
7	30	88.2	1609	1 MMHUR2	laminin gamma-1 ch
8	29	85.3	308	2 A83142	probable ATP-bind
9	29	85.3	605	2 S67815	protein-tyrosine k
10	29	85.3	640	2 T09534	probable beta-fruc
11	29	85.3	764	2 JC4736	methvl-accepting c
12	29	85.3	1756	2 T07566	hypothetical prote
13	28	82.4	263	2 A54543	beta-lactamase (EC
14	28	82.4	298	2 S12579	cardonate dehydrat
15	28	82.4	578	2 DB7374	RsaA secretion sys
16	27	79.4	123	2 F71267	probable ribosomal
17	27	79.4	153	2 H82111	flagellin flag VC2
18	27	79.4	226	2 T43814	conserved hypotnet
19	27	79.4	283	2 G84321	hypothetical prote
20	27	79.4	323	2 G83461	hypothetical prote
21	27	79.4	349	2 S74439	iron(III) diclitrac
22	27	79.4	373	2 T38687	hypothetical prote
23	27	79.4	390	2 D70834	hypothetical prote
24	27	79.4	424	2 E87558	cytochrome P450 fa
25	27	79.4	508	2 G84339	phosphoglycerate m
26	27	79.4	707	2 JC2218	procollagen C-endo
27	27	79.4	1141	2 T20611	hypothetical prote
28	27	79.4	1245	2 G86404	probable P-glycopr
29	27	79.4	2174	2 E95965	hypothetical glycol

30	27	79.4	3396	1 A42551	genome polypeptide
31	26	76.5	60	2 PNO674	GTP-binding prote
32	26	76.5	125	2 C87560	hypothetical prote
33	26	76.5	126	2 G84215	hypothetical prote
34	26	76.5	152	2 A87679	conserved hypotnet
35	26	76.5	161	2 AB2583	conserved hypotnet
36	26	76.5	187	2 C71140	hypothetical prote
37	26	76.5	191	2 H97364	hypothetical prote
38	26	76.5	199	2 F83540	probable alanyl hyd
39	26	76.5	217	2 AD2683	Conserved hypotnet
40	26	76.5	234	2 T36369	response regulator
41	26	76.5	239	2 AB3193	conserved hypotnet
42	26	76.5	245	2 S63644	cytochrome-c oxida
43	26	76.5	252	2 AE1302	probable phosphopr
44	26	76.5	252	2 AE1674	probable phosphopr
45	26	76.5	253	2 B97465	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

KGHUGH histidine-rich glycoprotein precursor - human

N:Alternate names: HRG

C:Species: Homo sapiens (man)

C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000

C:Accession: A01287; S29669

R:Koida, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A:Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu

A:Reference number: A01287; MUID:86216149

A:Accession: A01287

A:Molecule type: mRNA

A:Residues: 1-525 <KOI>

A:Cross-references: GB:AB005803; NID:q2280513; PIDN:BAA21613.1; PID:q2280514

R:Hemis, B.; Havelaar, A.; Kluff, C.

submitted to the EMBL Data Library, October 1991

A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly

A:Reference number: S29669

A:Accession: S29669

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 214-247 <HEN>

A:Cross-references: EMBL:217218; NID:q32453; PIDN:CAA78925.1; PID:q32454

C:Comment: Although its physiological function is not yet known, HRG does bind heme, din, and the lysine-binding site of plasminogen. On the basis of its homology with HM

C:Comment: The amino half of this protein is homologous to the first two cystatin-11k

C:Comment: In addition to having a high histidine and proline content, this protein h

e-rich region.

C:Genetics:

A:Gene: GDB:HRG

A:Cross-references: GDB:120055; OMIM:142640

A:Map position: 3q27-3q27

C:Superfamily: histidine-rich glycoprotein; cystatin homology

C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F:19-131/Domain: cystatin homology <CV1>

F:140-246/Domain: cystatin homology <CV2>

F:276-321/Region: proline-rich

F:348-437/Region: histidine-rich

F:351-497/Region: proline-rich

F:63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 100.0%; Score 34; DB 1; Length 525;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
|||||  
Db 173 ASFRVDR 179

## RESULT 2

B96806  
hypothetical protein TSM16.24 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B96806  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marshall,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: B96806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-383 <STO>  
A:Cross-references: GB:AE005173; NID:g6382509; PIDN:AAF07795.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: TSM16.24  
A:Map position: 1

Query Match 91.2%; Score 31; DB 2; Length 383;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
|||||  
Db 250 SSFRVDR 256

## RESULT 3

C87703  
conserved hypothetical protein CC3661 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: C87703  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Debay, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: C87703  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-179 <STO>  
A:Cross-references: GB:AE005673; NID:g13425419; PIDN:AAK25623.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC3661

Query Match 88.2%; Score 30; DB 2; Length 179;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
|||||  
Db 149 SFRVDR 154

## RESULT 4

AD3445

hypothetical protein al1516 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AD2445  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2445  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-568 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA076815.1; PID:g17134254; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: al1516

Query Match 88.2%; Score 30; DB 2; Length 588;  
Best Local Similarity 85.7%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
|||||  
Db 94 AKFRVDR 100

## RESULT 5

T00331  
hypothetical protein KIAA0555 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00331  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
A:Reference number: Z14086; MUID:98290545  
A:Accession: T00331  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-799 <NAG>  
A:Cross-references: EMBL:AB011127; NID:g3043633; PIDN:BAA25481.1; PID:g3043634  
A:Experimental source: brain; clone HH0882  
C:Genetics:  
A:Note: KIAA0555

Query Match 88.2%; Score 30; DB 2; Length 799;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
|||||  
Db 450 ASFRVDR 456

## RESULT 6

MMSB2  
laminin gamma-1 chain precursor - mouse  
N:Alternate names: laminin chain B2  
C:Species: Mus musculus (house mouse)  
C:Date: 28-Feb-1986 #sequence\_revision 30-Jun-1991 #text\_change 10-Dec-1999  
C:Accession: A28469; A27729; A28082; S02680; S05327; S02037; A02870; S13544; S14552  
R:Sasaki, M.; Yamada, Y.  
J. Biol. Chem. 262, 17111-17117, 1987  
A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.  
A:Reference number: A28469; MUID:88059118  
A:Accession: A28469  
A:Molecule type: mRNA  
A:Residues: 1-1607 <SAS>  
A:Cross-references: EMBL:J03484; NID:g198694; PIDN:AAA39405.1; PID:g293688

R:Durkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.  
 Biochemistry 27, 5198-5204, 1988  
 A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin B1.  
 A:Reference number: A27729; MUID:89000737  
 A:Accession: A27729  
 A:Molecule type: mRNA  
 A:Residues: 1-263, 'D', 265-336, 'C', 338-446, 'P', 449-661, 'S', 663-885, 887-1155, 1157-1433, 'A'  
 A:Cross-references: EMBL:J02930; NID:9198702; PDB:AAA39408.1; PID:9293691  
 A>Note: the authors translated the codon TAT for residue 544 as Asp and GCG for residue  
 U. Biol. Chem. 263, 8384-8389, 1988  
 R:Ogawa, K.; Birdelo, P.D.; Sasaki, M.; Yamada, Y.  
 A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active in  
 A:Reference number: A28082; MUID:88228071  
 A:Accession: A28082  
 A:Molecule type: DNA  
 A:Residues: 1-215, 'A', 217-239 <OGA>  
 A:Cross-references: EMBL:J03749; NID:9198704; PDB:AAA39409.1; PID:9554184  
 R:Fujikawa, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.  
 Biochem. J. 252, 453-461, 1988  
 A:Title: Structure and distribution of N-linked oligosaccharide chains on various domain  
 A:Reference number: S02678; MUID:88326259  
 A:Accession: S02680  
 A:Molecule type: Protein  
 A:Residues: 227-238 <FUY>  
 R:Hartl, L.; Oberbauer, I.; Deutzmann, R.  
 Eur. J. Biochem. 173, 629-635, 1988  
 A:Title: The N terminus of laminin A chain is homologous to the B chains.  
 A:Reference number: S00624; MUID:88225080  
 A:Accession: S05327  
 A:Molecule type: Protein  
 A:Residues: 1362-1377, 'X', 1379-1392, 'X', 1394-1406 <DEU>  
 R:Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.  
 EMBO J. 3, 2355-2362, 1984  
 A:Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil  
 A:Reference number: A02870; MUID:85051302  
 A:Accession: A02870  
 A:Molecule type: mRNA  
 A:Residues: 1391-1474, 'K', 1476-1575, 'N', 1577-1607 <BAR>  
 A:Cross-references: EMBL:X05211; NID:952862; PDB:CAA28838.1; PID:9817975  
 R:Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.  
 EMBO J. 4, 309-316, 1985  
 A:Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.  
 A:Reference number: S13543; MUID:85257455  
 A:Accession: S13544  
 A:Molecule type: Protein  
 A:Residues: 1506-1523, 'X', 1525 <PAU>  
 R:Olson, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,  
 Lab. Invest. 60, 772-782, 1989  
 A:Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 ch  
 A:Reference number: A34961; MUID:89280632  
 A:Accession: S14552  
 A:Molecule type: protein  
 A:Residues: 881-912;1022-1034;1364-1377;1379-1392;1394-1409;1506-1525;1593-1606 <OLS>  
 A:Genetics:  
 A:Gene: Lamb-2  
 A:Map position: 1  
 A:introns: 138/1; 239/3  
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin  
 C:Function:  
 A:Description: Interact with cells and with other basement membrane proteins to promote  
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology  
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular  
 F:1-33/Domain: signal sequence #status predicted <SIG>  
 F:34-1607/Product: laminin gamma-1 chain #status predicted <MAT>  
 F:34-283/Domain: VI <DOM6>  
 F:284-502/Domain: V <DOM5>

F:284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>  
 F:340-393/Domain: laminin-type EGF-like homology <LE02>  
 F:396-440/Domain: laminin-type EGF-like homology <LE03>  
 F:443-490/Domain: laminin-type EGF-like homology <LE04>  
 F:493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>  
 F:503-687/Domain: IV <DOM4>  
 F:688-1032/Domain: III <DOM3>  
 F:688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>  
 F:722-768/Domain: laminin-type EGF-like homology <LE07>  
 F:771-823/Domain: laminin-type EGF-like homology <LE08>  
 F:826-879/Domain: laminin-type EGF-like homology <LE09>  
 F:882-930/Domain: laminin-type EGF-like homology <LE10>  
 F:933-978/Domain: laminin-type EGF-like homology <LE11>  
 F:981-1026/Domain: laminin-type EGF-like homology <LE12>  
 F:1033-1607/Domain: II/I <DOM2>  
 F:1033-1607/Region: heptad repeats  
 F:38-48/Disulfide bonds: #status predicted  
 F:58-132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/binding site: carbohydrate (F:1029,1032/Disulfide bonds: interchain #status predicted  
 F:1378,1393/binding site: carbohydrate (Asn) (covariant) #status experimental  
 F:1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match 88.2%; Score 30; DB 1; Length 1607;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 DB 578 SFRVDR 583  
 |||||

RESULT 7  
 MMHDB2  
 laminin gamma-1 chain precursor - human  
 N:Alternate names: laminin chain B2  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1991 #sequence, revision 30-Jun-1991 #text, change 10-Dec-1999  
 C:Accession: S13548; A28158; S13549; B34961; S14664; S23567  
 R:Kallunki, T.; Ikonen, U.; Chow, L.T.; Kallunki, P.; Tryggvason, K.  
 J. Biol. Chem. 266, 221-228, 1991  
 A:Title: Structure of the human laminin B2 chain gene reveals extensive divergence fr  
 A:Reference number: S13548; MUID:91093128  
 A:Accession: S13548  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1609 <KAL>  
 A:Cross-references: GB:M55217; NID:9186937  
 A>Note: the nucleotide sequence was submitted to GenBank, February 1991  
 R:Piikarainen, T.; Kallunki, T.; Tryggvason, K.  
 J. Biol. Chem. 263, 6751-6758, 1988  
 A:Title: Human laminin B2 chain. Comparison of the complete amino acid sequence with  
 A:Reference number: A28158; MUID:88198245  
 A:Accession: A28158  
 A:Molecule type: mRNA  
 A:Residues: 1-211, 'I', 213-1609 <PIK>  
 A:Cross-references: EMBL:J03202; NID:9186916; PDB:AAA59488.1; PID:9307107  
 R:Fukushima, Y.; Piikarainen, T.; Kallunki, T.; Eddy, R.L.; Byers, M.G.; Haley, L.L.;  
 Cytogenet. Cell Genet. 48, 137-141, 1988  
 A:Title: Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of the gen  
 A:Reference number: S13549; MUID:89169663  
 A:Accession: S13549  
 A:Molecule type: mRNA  
 A:Residues: 1393-1609 <FUK>  
 A:Cross-references: EMBL:M27654; NID:9186923; PDB:AAA59489.1; PID:9186924  
 R:Olson, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,  
 Lab. Invest. 60, 772-782, 1989  
 A:Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2  
 A:Reference number: A34961; MUID:89280632  
 A:Accession: B34961  
 A:Molecule type: mRNA  
 A:Residues: 868-1551, 'N', 1553-1609 <OLS>  
 R:Santos, C.L.S.; Sabbaga, J.; Brentani, R.

DNA Seq. 1, 275-277, 1991  
 A>Title: Differences in human laminin B2 sequences.  
 A:Reference number: S14664; MUID:92216129  
 A:Accession: S14664  
 A:Molecule type: mRNA  
 A:Residues: 1282-1609 <S>  
 A:Cross-references: EMBL:X13939; NID:g34237; PID:CAA32122.1; PID:g34238  
 R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Pikkariainen, T.; Tryggvason, K.  
 In Extracellular Matrix Genes, Sandell L.J., eds., pp. 175-193, Academic F  
 A>Title: Genes for the human laminin B1 and B2 chains.  
 A:Reference number: S23566  
 A:Accession: S23567  
 A:Molecule type: DNA  
 A:Residues: 801-1461, 'R', 1483-1609 <V>  
 A>Note: mRNA was also sequenced  
 C:Genetics:  
 A:Gene: GDB:LAMC1; LAMB2  
 A:Cross-references: GDB:120136; OMIM:150290  
 A:Map position: 1q31-q31  
 A:Introns: 140/1; 241/3; 285/2; 341/1; 404/1; 443/2; 476/2; 522/1; 563/1; 626/2; 664/1;  
 /3; 1525/1  
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin  
 C:Function:  
 A:Description: Interact with cells and with other basement membrane proteins to promote  
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology  
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular  
 F:1-33/Domain: signal sequence #status predicted <SIG>  
 F:34-1609/Product: laminin gamma-1 chain #status predicted <MAT>  
 F:34-285/Domain: VI <DOM6>  
 F:286-504/Domain: V <DOM5>  
 F:286-339/Domain: laminin-type EGF-like homology <LE01>  
 F:342-395/Domain: laminin-type EGF-like homology <LE02>  
 F:398-442/Domain: laminin-type EGF-like homology <LE03>  
 F:445-492/Domain: laminin-type EGF-like homology <LE04>  
 F:495-504/Domain: laminin-type EGF-like homology #status atypical <LE05>  
 F:505-689/Domain: IV <DOM4>  
 F:690-1034/Domain: III <DOM3>  
 F:690-722/Domain: laminin-type EGF-like homology #status atypical <LE06>  
 F:724-770/Domain: laminin-type EGF-like homology <LE07>  
 F:773-823/Domain: laminin-type EGF-like homology <LE08>  
 F:828-881/Domain: laminin-type EGF-like homology <LE09>  
 F:884-933/Domain: laminin-type EGF-like homology <LE10>  
 F:935-980/Domain: laminin-type EGF-like homology <LE11>  
 F:983-1028/Domain: laminin-type EGF-like homology <LE12>  
 F:1035-1609/Domain: II/I <DOM1>  
 F:1035-1609/Region: heptad repeats  
 F:40-50/Disulfide bonds: #status predicted  
 F:60,134,576,650,1022,1107,1161,1175,1205,1223,1241,1380,1395,1439/Binding site: carboxy  
 F:1031,1034,1600/Disulfide bonds: interchain #status predicted

Query Match 88.2%; Score 30; DB 1; Length 1609;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFRVDR 7  
 |||||  
 Db 580 SFRVDR 585

RESULT 8  
 A83142  
 Problem: ATP-binding component of ABC transporter PA4037 [Imported] - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: A83142  
 R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: A83142

A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-308 <S>  
 A:Cross-references: GB:AE004820; GB:AE004091; NID:g9950223; PID:AA07424.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4037  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 85.3%; Score 29; DB 2; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFRVDR 7  
 |||||  
 Db 21 SFRVDR 26

RESULT 9  
 S67815  
 protein-tyrosine kinase (EC 2.7.1.112) GCTR - Geodia cydonium  
 N:Alternate names: receptor tyrosine kinase  
 C:Species: Geodia cydonium  
 C>Date: 24-Aug-1996 #sequence\_revision 13-Mar-1997 #text\_change 24-Sep-1999  
 C:Accession: S67815; S67946; S32150  
 R:Mueller, W.E.G.  
 submitted to the EMBL Data Library, December 1995  
 A:Reference number: S67815  
 A:Accession: S67815  
 A:Molecule type: mRNA  
 A:Residues: 1-605 <MUE>  
 A:Cross-references: EMBL:X72622; NID:g1103390; PID:CAA51198.1; PID:e212400; PID:g110  
 A>Note: this is a revision to the sequence from reference S67946  
 R:Schaeke, H.; Schroeder, H.C.; Gamulin, V.; Rinkewich, B.; Mueller, I.M.; Mueller,  
 M. Membr. Biol. 11, 101-107, 1994  
 A>Title: Molecular cloning of a tyrosine kinase gene from the marine sponge Geodia cy  
 A:Reference number: S67946; MUID:95004895  
 A:Accession: S67946  
 A:Molecule type: mRNA  
 A:Residues: 1-605 <MUE>  
 A:Cross-references: EMBL:X72622  
 A>Note: this sequence has been revised in reference S67815  
 R:Pfeifer, K.; Schaeke, H.; Schroeder, H.C.; Gamulin, V.; Mueller, I.M.; Mueller, W.E  
 submitted to the EMBL Data Library, March 1993  
 A:Description: cDNA cloning and function of a receptor tyrosine kinase gene from the  
 A:Reference number: S32150  
 A:Accession: S32150  
 A:Molecule type: mRNA  
 A:Residues: 58-79, 'E', 81-227, 'TL', 231, 'QGPSE', 237-248, 'RR', 252-255, 'L', 257-274, 'SRQ',  
 A:Cross-references: EMBL:X72622  
 A>Note: this sequence has been revised in reference S67946  
 C:Genetics:  
 A:Gene: GCTR  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
 C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase  
 F:344-605/Domain: protein kinase homology <KIN>  
 F:352-360/Region: protein kinase ATP-binding motif

Query Match 85.3%; Score 29; DB 2; Length 605;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVD 6  
 |||||  
 Db 62 ASFRVD 67

RESULT 10  
 T09534  
 probable beta-fructofuranosidase (EC 3.2.1.26) - pepper  
 N:Alternate names: acid beta-fructosidase

C:Species: Capsicum annuum (pepper)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 C:Accession: T09534  
 R:Choi, D.; Kim, S.; Lee, K.W.  
 submitted to the EMBL Data Library, January 1997  
 A:Reference number: Z16722  
 A:Accession: T09534  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-640 <CHO>  
 A:Cross-References: EMBL:U87849; NID:g1857713; PID:g1857714  
 C:Function:  
 A:Description: catalyzes the hydrolyzation of terminal non-reducing beta-D-fructofuranos  
 C:Superfamily: beta-fructofuranosidase  
 C:Keywords: glycosidase; hydrolase

Query Match 85.3%; Score 29; DB 2; Length 640;  
 Best Local Similarity 85.7%; Pred. No. 67;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
 :|||:|  
 DB 471 ASFEVDR 477

RESULT 11  
 Jc4736  
 methyl-accepting chemotaxis protein B - Rhodobacter capsulatus  
 C:Species: Rhodobacter capsulatus  
 C:Date: 10-May-1996 #sequence\_revision 16-Aug-1996 #text\_change 08-Oct-1999  
 C:Accession: Jc4736  
 R:Michotey, V.; Tousseint, B.; Richard, P.; Vignais, P.M.  
 Gene 170, 73-76, 1996  
 A:Title: Characterisation of the mcpA and mcpB genes capable of encoding methyl-acceptin  
 A:Reference number: Jc4735; MUID:96200858  
 A:Accession: Jc4736  
 A:Molecule type: DNA  
 A:Residues: 1-764 <MIC>  
 A:Cross-References: GB:I48927; NID:g1066849; PIDN:AA05215.1; PID:g1066851  
 C:Genetics:  
 A:Gene: mcpB  
 C:Keywords: transmembrane protein  
 F:1-22/Domain: hydrophobic #status predicted <HYD>  
 F:266-295/Domain: hydrophobic #status predicted <HYD>

Query Match 85.3%; Score 29; DB 2; Length 764;  
 Best Local Similarity 71.4%; Pred. No. 80;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
 :|||:|  
 DB 55 AKFRIDR 61

RESULT 12  
 T07566  
 hypothetical protein 756 - Japanese black pine chloroplast  
 C:Species: chloroplast Pinus thunbergiana (Japanese black pine)  
 C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 18-Aug-2000  
 C:Accession: T07566  
 R:Wakusui, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugitara, M.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994  
 A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome  
 A:Reference number: Z16030; MUID:95024047  
 A:Accession: T07566  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1756 <MAK>  
 A:Cross-References: EMBL:D17510; NID:g529643; PIDN:BA04442.1; PID:g1262727  
 C:Genetics:  
 A:Genome: chloroplast

C:Keywords: chloroplast

Query Match 85.3%; Score 29; DB 2; Length 1756;  
 Best Local Similarity 83.3%; Pred. No. 19+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFRVDR 7  
 :|||:|  
 DB 188 SFRIDR 193

RESULT 13  
 A54543  
 beta-lactamase (EC 3.5.2.6), cefotaxime-hydrolyzing - Klebsiella oxytoca (strain D483  
 C:Species: Klebsiella oxytoca  
 C:Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 03-May-1996  
 C:Accession: A54543  
 R:Reynaud, A.; Peduzzi, J.; Barthelemy, M.; Labie, R.  
 FEMS Microbiol. Lett. 81, 185-192, 1991  
 A:Title: Cefotaxime-hydrolysing activity of the beta-lactamase of Klebsiella oxytoca  
 A:Reference number: A54543  
 A:Accession: A54543  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-263 <REV>  
 A:Note: parts of this sequence were determined by composition rather than by peptide  
 C:Superfamily: beta-lactamase I  
 C:Keywords: antibiotic resistance; hydrolase

Query Match 82.4%; Score 28; DB 2; Length 263;  
 Best Local Similarity 71.4%; Pred. No. 47;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
 :|||:|  
 DB 133 AFRIDR 139

RESULT 14  
 S12579  
 carnate dehydratase (EC 4.2.1.1), hepatic - mouse  
 N:Alternate names: carbonic anhydrase  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 22-Jun-1999  
 C:Accession: S12579  
 R:Amor-Gueret, M.; Levi-Strauss, M.  
 Nucleic Acids Res. 18, 1646, 1990  
 A:Title: Nucleotide and derived amino-acid sequence of a cDNA encoding a new mouse ca  
 A:Reference number: S12579; MUID:90221910  
 A:Accession: S12579  
 A:Molecule type: mRNA  
 A:Residues: 1-298 <AMO>  
 A:Cross-References: EMBL:X51971; NID:950284; PIDN:CA36233.1; PID:950285  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; liver; mitochondrion; zinc  
 F:35-289/Domain: carbonic anhydrase homology <CAH>

Query Match 82.4%; Score 28; DB 2; Length 298;  
 Best Local Similarity 71.4%; Pred. No. 54;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
 :|||:|  
 DB 287 SSFRIDR 293

RESULT 15  
 D87374  
 RsaA secretion system, ATP-binding protein RsaD [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence-revision 20-Apr-2001 #text-change 20-Apr-2001  
C;Accession: D87374  
R;Mierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: D87374  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1578 <STO>  
A;Cross-references: GB:AE005673; NID:q13422298; PIDN:AAK22992.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC1008

Query Match 82.4%; Score 28; DB 2; Length 578;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVD 6  
|111:1  
Db 350 ASFRID 355

Search completed: May 24, 2002, 16:50:17  
Job time: 322 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:49:03 ; Search time 49.37 Seconds  
(without alignments)  
3.463 Million cell updates/sec

Title: US-09-730-379E-6

Perfect score: 34

Sequence: 1 ASFRVDR 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfill1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description
	1	29	85.3	24	1	US-08-105-416-24	Sequence 24, Appl
	2	29	85.3	24	2	US-08-473-656A-24	Sequence 24, Appl
	3	29	85.3	24	3	US-08-483-931B-24	Sequence 24, Appl
	4	26	76.5	421	2	US-08-659-984A-1	Sequence 1, Appl
	5	26	76.5	421	4	US-08-660-531-1	Sequence 1, Appl
	6	26	76.5	444	2	US-08-659-984A-5	Sequence 5, Appl
	7	26	76.5	444	4	US-08-660-531-5	Sequence 5, Appl
	8	26	76.5	635	1	US-08-245-809-1	Sequence 1, Appl
	9	26	76.5	636	1	US-08-296-624-2	Sequence 1, Appl
	10	26	76.5	803	1	US-08-062-368-4	Sequence 2, Appl
	11	26	76.5	972	3	US-08-335-844A-23	Sequence 23, Appl
	12	26	76.5	972	3	US-08-335-844A-24	Sequence 24, Appl
	13	26	76.5	977	3	US-08-335-844A-22	Sequence 22, Appl
	14	25	73.5	11	4	US-09-028-937-41	Sequence 41, Appl
	15	25	73.5	17	1	US-07-976-872B-1	Sequence 1, Appl
	16	25	73.5	17	1	US-07-976-872B-2	Sequence 2, Appl
	17	25	73.5	17	1	US-07-976-872B-3	Sequence 3, Appl
	18	25	73.5	17	1	US-07-976-872B-4	Sequence 4, Appl
	19	25	73.5	17	1	US-08-105-416-12	Sequence 12, Appl
	20	25	73.5	17	1	US-08-105-416-13	Sequence 13, Appl
	21	25	73.5	17	1	US-08-105-416-15	Sequence 15, Appl
	22	25	73.5	17	1	US-08-105-416-17	Sequence 17, Appl
	23	25	73.5	17	1	US-08-105-416-18	Sequence 18, Appl
	24	25	73.5	17	1	US-08-105-416-19	Sequence 19, Appl
	25	25	73.5	17	1	US-08-105-416-20	Sequence 20, Appl
	26	25	73.5	17	1	US-08-105-416-28	Sequence 28, Appl
	27	25	73.5	17	1	US-08-105-416-37	Sequence 37, Appl

28	25	73.5	17	1	US-08-105-416-38	Sequence 38, Appl
29	25	73.5	17	1	US-08-105-416-39	Sequence 39, Appl
30	25	73.5	17	1	US-08-105-416-40	Sequence 40, Appl
31	25	73.5	17	2	US-08-473-656A-12	Sequence 12, Appl
32	25	73.5	17	2	US-08-473-656A-13	Sequence 13, Appl
33	25	73.5	17	2	US-08-473-656A-15	Sequence 15, Appl
34	25	73.5	17	2	US-08-473-656A-17	Sequence 17, Appl
35	25	73.5	17	2	US-08-473-656A-18	Sequence 18, Appl
36	25	73.5	17	2	US-08-473-656A-19	Sequence 19, Appl
37	25	73.5	17	2	US-08-473-656A-20	Sequence 20, Appl
38	25	73.5	17	2	US-08-473-656A-28	Sequence 28, Appl
39	25	73.5	17	2	US-08-473-656A-37	Sequence 37, Appl
40	25	73.5	17	2	US-08-473-656A-38	Sequence 38, Appl
41	25	73.5	17	2	US-08-473-656A-39	Sequence 39, Appl
42	25	73.5	17	2	US-08-473-656A-40	Sequence 40, Appl
43	25	73.5	17	3	US-08-483-931B-12	Sequence 12, Appl
44	25	73.5	17	3	US-08-483-931B-13	Sequence 13, Appl
45	25	73.5	17	3	US-08-483-931B-15	Sequence 15, Appl

#### ALIGNMENTS

RESULT 1  
US-08-105-416-24  
Sequence 24, Application US/08105416  
Patent No. 5639958  
GENERAL INFORMATION:  
APPLICANT: Olsson, Lennart S  
APPLICANT: Goodenow, Robert S  
TITLE OF INVENTION: Class I MHC Modulation of Surface  
TITLE OF INVENTION: Receptor Activity  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Flehr, Hohnbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/105, 416  
FILING DATE: 12-AUG-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Berttram I.  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A5115-4/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
SPRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-105-416-24

Query Match 85.3%; Score 29; DB 1; Length 24;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ASFRVD 6  
||||||

Db 11 ASFRVD 16

## RESULT 2

US-08-473-656A-24

; Sequence 24, Application US/08473656A

; Patent No. 5853999

## GENERAL INFORMATION:

APPLICANT: Olsson, Lennart

APPLICANT: Goodenow, Robert S

APPLICANT: Goldstein, Avram

TITLE OF INVENTION: Class I MHC Modulation of Surface

TITLE OF INVENTION: Receptor Activity

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Flehr, Hohbach, Test, Albritton &amp; Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: US

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/473,656A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/105,416

FILING DATE: 12-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bertlam I.

REGISTRATION NUMBER: 20015

REFERENCE/DOCKET NUMBER: A55115-4/BIR

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-473-656A-24

Query Match 85.3%; Score 29; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVD 6

Db 11 ASFRVD 16

## RESULT 3

US-08-483-931B-24

; Sequence 24, Application US/08483931B

; Patent No. 6028171

## GENERAL INFORMATION:

APPLICANT: Olsson, Lennart

APPLICANT: Goodenow, Robert S.

APPLICANT: Goldstein, Avram

TITLE OF INVENTION: Methods of Screening for Agents

TITLE OF INVENTION: That Inhibit Interaction Between MHC Class I

TITLE OF INVENTION: Antigens and a Cell Surface Receptor

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,931B

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/028,241

FILING DATE: 20-MAR-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/323,565

FILING DATE: 14-MAR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/649,471

FILING DATE: 01-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,184

FILING DATE: 03-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/105,416

FILING DATE: 12-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: Hunter, Tom

REGISTRATION NUMBER: 38,498

REFERENCE/DOCKET NUMBER: 02307E-07225005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-483-931B-24

Query Match 85.3%; Score 29; DB 3; Length 24;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFRVD 6

Db 11 ASFRVD 16

## RESULT 4

US-08-659-984A-1

; Sequence 1, Application US/08659984A

; Patent No. 5942400

## GENERAL INFORMATION:

APPLICANT: Anderson, John P.

APPLICANT: Jacobson-Croak, Kirsten L.

TITLE OF INVENTION: Assays for Detecting Beta-Secretase

TITLE OF INVENTION: Inhibition

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Ctr., 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 421 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-1

Query Match 76.5%; Score 26; DB 2; Length 421;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FRVDR 7  
|||||  
DB 176 FRVDR 180

RESULT 5  
US-08-660-531-1  
Sequence 1, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 421 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-660-531-1

Query Match 76.5%; Score 26; DB 4; Length 421;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FRVDR 7  
|||||  
DB 176 FRVDR 180

RESULT 6  
US-08-659-984A-5  
Sequence 5, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-5

Query Match 76.5%; Score 26; DB 2; Length 444;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FRVDR 7

Db 199 FRVDR 203

## RESULT 7

US-08-660-531-5  
; Sequence 5, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-00221005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ. ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-660-531-5

Query Match 76.5%; Score 26; DB 4; Length 444;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FRVDR 7  
Db 199 FRVDR 203

## RESULT 8

US-08-245-809-1  
; Sequence 1, Application US/08245809  
; Patent No. 5665579  
; GENERAL INFORMATION:  
; APPLICANT: Fitzmaurice, Leona  
; APPLICANT: Mirkov, Erik  
; APPLICANT: Butler, William O.  
; APPLICANT: Dickinson, Craig D.  
; APPLICANT: Elliott, Kathryn J.  
; APPLICANT: Konno, Yoshitiro  
; TITLE OF INVENTION: NOVEL INVERTASE GENE(S) AND USES THEREOF  
; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitch, Even, Tabin & Flannery  
; STREET: 135 South LaSalle Street, Suite 900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: U.S.A.  
; ZIP: 60603-4277

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/245,809  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/771,331  
; FILING DATE: 03-OCT-1991  
; APPLICATION NUMBER: US 07/660,344  
; FILING DATE: 22-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 51514  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-552-1311  
; INFORMATION FOR SEQ. ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-245-809-1

Query Match 76.5%; Score 26; DB 1; Length 635;  
Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASFRVDR 7  
Db 465 ASFEVDK 471

## RESULT 9

US-08-296-624-2  
; Sequence 2, Application US/08296624  
; Patent No. 5658773  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Klann, Ellen  
; TITLE OF INVENTION: Tomato Acid Invertase Gene  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Weber  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,624  
; FILING DATE: 26-AUG-1994  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/770,970  
; FILING DATE: 07-OCT-1991

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 2307E-036510US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-296-624-2

Query Match          76.5%; Score 26; DB 1; Length 636;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7
Db 466 ASFEVDK 472

RESULT 10
US-08-062-368-4
; Sequence 4, Application US/08062368
; Patent No. 5491086
; GENERAL INFORMATION:
; APPLICANT: Gelland, David H.
; APPLICANT: Wang, Alice M.
; TITLE OF INVENTION: Purified Thermostable Nucleic Acid
; TITLE OF INVENTION: Polymerases Enzyme From Pyrodicticum Species
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,368
; FILING DATE: 19930514
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D., Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: 8584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-062-368-4

Query Match          76.5%; Score 26; DB 1; Length 803;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FRVDR 7

; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Trevor Stanley
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-335-844A-23

Query Match          76.5%; Score 26; DB 3; Length 972;
Best Local Similarity 66.7%; Pred. No. 5.6e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7
Db 448 SFRIDK 453

RESULT 12
US-08-335-844A-24
; Sequence 24, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MUNN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,844A
; FILING DATE: 09-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9209936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-335-844A-23

Query Match          76.5%; Score 26; DB 3; Length 972;
Best Local Similarity 66.7%; Pred. No. 5.6e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7
Db 448 SFRIDK 453

RESULT 12
US-08-335-844A-24
; Sequence 24, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
```

APPLICANT: MUNN, EDWARD ALBERT  
APPLICANT: KNOX, DAVID PATRICK  
APPLICANT: OLIVER, JOANNA JANE  
APPLICANT: NEWTON, SUSAN ELIZABETH  
TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF  
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
STREET: Suite 701-E, 555 Thirteenth St., N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/335,844A  
FILING DATE: 09-JAN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB PCT/GB93/00943  
FILING DATE: 06-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9209936  
FILING DATE: 08-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, Barbara W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 1181-223A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 972 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-335-844A-24

Query Match 76.5%; Score 26; DB 3; Length 972;  
Best Local Similarity 66.7%; Pred. No. 5.6e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SFRVDR 7  
DB 448 SFRIDK 453

RESULT 13  
US-08-335-844A-22  
Sequence 22, Application US/08335844A  
Patent No. 6066503  
GENERAL INFORMATION:  
APPLICANT: GRAHAM, MARGARET  
APPLICANT: SMITH, TREVOR STANLEY  
APPLICANT: MUNN, EDWARD ALBERT  
APPLICANT: KNOX, DAVID PATRICK  
APPLICANT: OLIVER, JOANNA JANE  
APPLICANT: NEWTON, SUSAN ELIZABETH  
TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF  
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
STREET: Suite 701-E, 555 Thirteenth St., N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/028,937  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/788,820  
FILING DATE: 23-JAN-1997

CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/335,844A  
FILING DATE: 09-JAN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB PCT/GB93/00943  
FILING DATE: 06-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9209936  
FILING DATE: 08-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, Barbara W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 1181-223A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 977 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-335-844A-22

Query Match 76.5%; Score 26; DB 3; Length 977;  
Best Local Similarity 66.7%; Pred. No. 5.6e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SFRVDR 7  
DB 455 SFRIDK 460

RESULT 14  
US-09-028-937-41  
Sequence 41, Application US/09028937  
Patent No. 6333031  
GENERAL INFORMATION:  
APPLICANT: Olsson, Lennart  
APPLICANT: Naranda, Tatjana  
TITLE OF INVENTION: Receptor Derived Peptides As Modulators  
TITLE OF INVENTION: Of Receptor Activity  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flenz, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/028,937  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/788,820  
FILING DATE: 23-JAN-1997

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/701,382  
FILING DATE: 22-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/612,999  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SILVA, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63139-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 949-8711  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-028-937-41

Query Match 73.5%; Score 25; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVD 6  
IIIII  
DB 1 SFRVD 5

RESULT 15  
US-07-976-872B-1  
Sequence 1, Application US/07976872B  
Patent No. 5516642  
GENERAL INFORMATION:  
APPLICANT: Mepelli, Claudio  
TITLE OF INVENTION: POLYPEPTIDES DERIVED FROM MAJOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burton Rodney  
STREET: P.O. Box 4000  
CITY: Princeton  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08543-4000  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/976,872B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaul, Timothy J.  
REGISTRATION NUMBER: 33,111  
REFERENCE/DOCKET NUMBER: KX41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 252-5901  
TELEFAX: (609) 252-4526  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-976-872B-1

Query Match 73.5%; Score 25; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVD 6  
IIIII  
DB 5 SFRVD 9

Search completed: May 24, 2002, 16:49:03  
Job time: 288 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:07 ; Search time 136.41 Seconds  
(without alignments)  
5.700 Million cell updates/sec

Title: US-09-730-379E-6  
Perfect score: 34  
Sequence: 1 ASFRVDR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	30	88.2	14 22 AAB56634	Human SNP related
2	30	88.2	91 22 AAU52904	Protonibacterium
3	30	88.2	139 10 AAP91214	Domain IV of mouse
4	30	88.2	140 20 AAY08449	Soybean Crml homol
5	30	88.2	324 22 AAG90292	C glutamicum prote
6	30	88.2	1572 21 AAB19806	Mouse laminin 2 ma
7	30	88.2	1572 21 AAB48455	Mouse laminin 8 po
8	30	88.2	1576 21 AAB48453	Human laminin 2 po
9	30	88.2	1584 21 AAB19804	Human laminin 2 ga
10	30	88.2	1584 21 AAB19804	Human laminin 2 ga
11	30	88.2	1605 21 AAB19805	Mouse laminin 2 ga

12	30	88.2	1605 21 AAB48454	Mouse laminin 8 po
13	30	88.2	1607 19 AAW50897	Mouse laminin G1 C
14	30	88.2	1609 19 AAW50898	Human laminin G1 C
15	30	88.2	1609 21 AAB19801	Human laminin 2 ga
16	30	88.2	1609 21 AAB48452	Human laminin 8 po
17	30	88.2	1617 21 AAB19803	Human laminin 2 ga
18	29	85.3	24 16 AAR71437	Human MHC 1 alpha
19	29	85.3	293 22 ABG08331	Novel human diagno
20	29	85.3	308 22 AAU33654	Pseudomonas aerugi
21	29	85.3	632 21 AAB26946	Wheat auxin transp
22	28	82.4	96 22 AAU64112	Protonibacterium
23	28	82.4	148 22 AA001880	Human polypeptide
24	28	82.4	555 21 AAB22786	Caulobacter cresce
25	28	82.4	634 22 ABB59653	Drosophila melanog
26	27	79.4	25 13 AAR20115	MHC Class I-deriva
27	27	79.4	35 22 AAB71820	Beta-hCG scram 1 p
28	27	79.4	138 21 AAB42110	Human ORFX ORF1874
29	27	79.4	188 22 AAU47717	Protonibacterium
30	27	79.4	219 21 AAY53872	Mature protein of
31	27	79.4	236 22 AAG90280	C glutamicum prote
32	27	79.4	244 21 AAY53871	A human Draln-deri
33	27	79.4	555 22 AAB31943	Amino acid sequenc
34	27	79.4	739 22 ABB67865	Drosophila melanog
35	27	79.4	763 27 ABB62815	Drosophila melanog
36	27	79.4	1274 22 AAB47329	FCFR6. Homo sapie
37	27	79.4	3392 22 AAE07980	Wild-type, virulen
38	27	79.4	3392 22 AAE07981	Attenuated, vaccin
39	27	79.4	3396 14 AAR43662	DENV1-5275/90 (ECAC
40	26	76.5	14 21 AAB09448	Hepatitis GB virus
41	26	76.5	59 22 AAU44304	Protonibacterium
42	26	76.5	100 21 AAB41057	Human ORFX ORF821
43	26	76.5	100 22 ABG05907	Novel human diagno
44	26	76.5	155 22 AAM25460	Human protein sequ
45	26	76.5	174 21 AAG36502	Arabidopsis thalia

## ALIGNMENTS

RESULT 1

AAB56634  
ID AAB56634 standard; Peptide: 14 AA.

AC AAB56634;  
XX 05-MAR-2002 (first entry)

DE Human SNP related amino acid sequence SEQ ID NO:1199.

XX Human: single nucleotide polymorphism; SNP; polymorphism; cytosstatic;  
XX immunosuppressive; antiinflammatory; neuroprotective; antimicrobial;  
XX autoimmune disease; inflammation; cancer; nervous system disease;  
XX Infection; polymorphic protein.

OS Homo sapiens.

XX WO200138586-A2.

PN 31-MAY-2001.

PD 22-NOV-2000; 2000WO-US32311.

PF 24-NOV-1999; 99US-0167383.

PR (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;  
PI WPI: 2001-355949/37.

DR Isolated human nucleic acids comprising one or more single nucleotide  
PT polymorphisms, useful for treating a subject suffering from a  
PT pathology, e.g. autoimmune diseases, ascribed to the presence of a

PT sequence polymorphism -  
 XX  
 PS Claim 1: Page 608; 674pp; English.  
 XX  
 CC ABL00010 to ABL01104 represent human nucleic acid oligonucleotides  
 CC comprising one or more single nucleotide polymorphisms (SNPs). ABB56531  
 CC to ABB56903 represent human peptides encoded by some of the SNP  
 CC oligonucleotides. The sequences from the present invention can have  
 CC immunosuppressive, cytostatic, antiinflammatory, neuroprotective and  
 CC antimicrobial activities. Nucleic acids, polypeptides, oligonucleotides  
 CC and antibodies from the present invention can be used for treating a  
 CC subject suffering from, at risk for, or suspected of, suffering from a  
 CC pathology ascribed to the presence of a sequence polymorphism. The  
 CC pathology may be autoimmune diseases, inflammation, cancer, diseases of  
 CC the nervous system, and infection by pathogenic microorganisms. The SNPs  
 CC are also useful for determining which forms of a characterised  
 CC polymorphism are present in individuals. The antibodies may be used in  
 CC the detection, quantitation and/or cellular or tissue localisation of a  
 CC polymorphic protein (e.g., for use in measuring levels of the  
 CC polymorphic protein within appropriate physiological samples).  
 CC  
 CC Sequence 14 AA;  
 SQ

Query Match 88.2%; Score 30; DB 22; Length 14;  
 Best Local Similarity 85.7%; Pred. No. 6.2;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
 |||||:|  
 Db 6 asfrver 12

## RESULT 2

AAU52904  
 ID AAU52904 standard; Protein; 91 AA.

AC AAU52904;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #13800.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI: 2001-616774/71.

DR N-ESDB: AAS59557.

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1: SEQ ID NO 14099; 1069pp; English.

XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 91 AA;  
 SQ

Query Match 88.2%; Score 30; DB 22; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
 |||||  
 Db 28 sfrvdr 33

## RESULT 3

AAAP91214  
 ID AAP91214 standard; protein; 139 AA.

AC AAP91214;

DT 01-JUN-1990 (first entry)

DE Domain IV of mouse laminin B2 chain.

KW Domain IV; mouse laminin A chain; mouse laminin B2 chain;

KW sequence homology; domain IVA; domain IVB.

OS Mouse.

PN US 7267564-A.

PD 28-FEB-1989.

PF 07-NOV-1988; 88US-0267564.

PR 07-NOV-1988; 88US-0267564.

PA (USSN ) US DEPT HEALTH & HUMAN.

PI Yamada Y, Sasaki M, Kleinman HK, Martin GR;

DR WPI: 1989-138175/18.

PT DNA encoding human laminin A chain, used in vector system -  
 PT and new synthetic peptide(s) with laminin-type biological  
 PT activity  
 XX  
 PS Fig 8; : 90pp; English.

XX There is significant homology (about 25 %) between it and domain IVA and  
 CC IVb of mouse laminin A chain. The program SEQHP aligns it to residues  
 CC 513 to 687 of domain IVA and residues 1,161 to 1,339 of domain IVB.

SQ Sequence 139 AA;

Query Match 88.2%; Score 30; DB 10; Length 139;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7  
| | | | |  
Db 35 sfrvdr 40

#### RESULT 4

AA08449  
ID AAY08449 standard; Protein; 140 AA.

AC AAY08449;

DT 27-JUL-1999 (first entry)

DE Soybean Crml homologue protein from clone srr3c.pk002.f10.

KW Padl; Crml; Jabl; AP-1 transcription factor activity; regulator; plant;

KW maize; soybean; wheat; rice; yeast; human; isolation; transgenic.

OS Glycine max.

PN WO9924574-A2.

PD 20-MAY-1999.

PF 04-NOV-1998; 98WO-US23487.

PR 07-NOV-1997; 97US-0064914.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Allen SM, Anderson SL, Hiltz WD, Kliney AJ, Miao G;

PI Morgante M, Odell JT, Sakai H;

DR WPI: 1999-327405/27.

DR N-PSDB: AAX55496.

PT Plant homologues of yeast Padl, Crml and human Jabl and related polynucleotides

PS Claim 6a; Page 46; 57pp; English.

CC This invention describes novel plant Padl, Crml or Jabl proteins which are capable of AP-1 transcription factor regulation. The proteins are thought to interact with transcription factors altering gene expression. CC The nucleic acid sequences of the invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Synthetic peptides of the proteins may be synthesized to generate antibodies that are useful for screening expression libraries. CC Transgenic plants may be produced using the nucleic acid sequences to alter the levels of Padl, Crml and Jabl present in the plants. Altering the levels of these proteins would alter the level of AP-1 transcription factor activity in the plants.

SQ Sequence 140 AA;

Query Match 88.2%; Score 30; DB 20; Length 140;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
| | | | |  
Db 109 asfrvdr 115

RESULT 5

AAG90292

ID AAG90292 standard; Protein; 324 AA.

AC AAG90292;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 4046.

KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

PA (KYOW ) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI: 2001-376931/40.

DR N-PSDB: AAH65511.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 4046; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a CC mutant of corynebacterium, measuring expression amount and CC analysing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, CC particularly L-lysine. The present sequence is a protein described CC in the exemplification of the invention. CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the CC European Patent Office.

SQ Sequence 324 AA;

Query Match 88.2%; Score 30; DB 22; Length 324;  
Best Local Similarity 71.4%; Pred. No. 12e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFRVDR 7  
: | | | | |  
Db 115 sfrvdr 121

#### RESULT 6

AA019806  
ID AAB19806 standard; Protein; 1572 AA.

AC AAB19806;

DT 05-MAR-2001 (first entry)

XX

DE	Mouse laminin 2 mature gamma-1 chain.
XX	
KW	Laminin 2; mouse; nerve regeneration; angiogenic; cell adhesion;
KM	degenerative muscle disorder; muscular dystrophy; cell therapy.
XX	
OS	Mus musculus.
PN	MO200066730-A2.
PD	09-NOV-2000.
XX	
PF	28-APR-2000; 2000WO-US11378.
XX	
PR	30-APR-1999; 99US-0131720.
XX	
PR	15-JUN-1999; 99US-0139198.
XX	
PR	12-JUL-1999; 99US-0143289.
XX	
PR	24-SEP-1999; 99US-0155945.
PA	(UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
PJ	
PI	Yurchenco P;
XX	
DR	WP1: 2000-687537/67.
XX	
DR	N-PSDB; AAA88906.
XX	
PT	Purified laminin 2 protein, useful for research and therapeutic
PT	purposes including peripheral nerve regeneration, treatment of
PT	degenerative muscle disorders, angiogenesis regulation, and ex vivo
PT	cell therapy -
PS	Citium 5; Page 302-306; 305pp; English.
XX	
CC	The present sequence is that of mouse laminin 2 gamma-1 chain mature
CC	protein. Laminin-2 is composed of alpha-2 (400 kDa), beta-1 (100
CC	kDa) and gamma-1 (100 kDa) chains. It is thought to be specifically
CC	required for stabilizing myotubes during skeletal muscle
CC	development, and for preventing apoptosis. Genetic defects in human
CC	laminin 2 structure or expression are associated with a major type
CC	of congenital muscular dystrophy. Laminin 2 is also thought to be
CC	important in Schwann cell/basal lamina interactions. The invention
CC	provides laminin 2 alpha-2, beta-1 and gamma-1 chain polypeptides
CC	(see AAB19791-806) and the polynucleotides encoding them (see
CC	AAA88891-906), methods for making recombinant laminin 2, cells that
CC	express recombinant laminin 2, and methods for using purified
CC	laminin 2 for research and therapeutic purposes including peripheral
CC	nerve regeneration, treatment of degenerative muscle disorders,
CC	angiogenesis regulation, promoting cell attachment and migration,
CC	ex vivo cell therapy, improving the take of grafts, improving the
CC	biocompatibility of medical devices and preparing improved culture
CC	devices and media.
XX	
SQ	Sequence 1572 AA;
QY	2 SFRVDR 7       545 sfrvdr 550
ID	Query Match 88.2%; Score 30; DB 21; Length 1572; Best Local Similarity 100.0%; Pred. No. 5.5e+02;
XX	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 7	
AAB48455	
ID	AAB48455 standard; Protein: 1572 AA.
XX	
AC	AAB48455;
XX	
DT	02-MAR-2001 (first entry)
XX	
DE	Mouse laminin 8 polypeptide, SEQ ID NO: 28.
XX	

KM		Mouse; laminin 8; neuroprotective; angiogenic; osteopathic;
KM		antiartherosclerotic; glycoprotein; mesenchymal tissue injury;
KM		vascular tissue injury; neural injury; angiogenesis regulation.
XX		
OS	Mus musculus.	
XX		
PN	WO2006732-AZ.	
XX		
PD	09-NOV-2000.	
PF	28-APR-2000; 2000WO-US11543.	
PR	30-APR-1999; 99US-0131720.	
PR	21-AUG-1999; 99US-0149738.	
PR	24-SEP-1999; 98US-0155945.	
XX	11-FEB-2000; 2000US-0182012.	
PA	(BIOS-) BIOSTRATUM INC.	
XX		
PI	Kortessmaa J, Tryggvason K;	
DR	WPI: 2000-687539/67.	
DR	N-PsDB; AACB3716.	
XX		
PT	Purified laminin 8 protein, useful for research and therapeutic purposes including peripheral nerve regeneration, treatment of degenerative muscle disorders, angiogenesis regulation, and ex vivo cell therapy -	
PS	Claim 5; Page 240-245; 245pp; English.	
CC	The present sequence is a laminin 8 polypeptide chain. Laminins are a family of heterotrimeric glycoproteins that function via binding interactions with neighbouring cell receptors and by forming laminin networks. They are signalling molecules which influence cellular function. Laminin 8 is useful for treating injuries to tissue of mesenchymal origin, such as bone, cartilage, tendon, and ligament, treating injuries to vascular tissue, promoting cell attachment and migration, ex vivo cell therapy, improving the biocompatibility of medical devices, and preparing improved cell culture devices and media. CC Laminin 8 is also useful for promoting re-endothelialisation at the site of vascular injuries, improving the take of grafts, improving the biocompatibility of medical devices, treating neural injuries (neural regeneration), regulating angiogenesis, and promoting cell attachment and migration..	
SQ	Sequence 1572 AA;	
	Query Match	88.2%; Score 30; DB 21; Length 1572;
	Best Local Similarity	100.0%; Pred. No. 5.5e+02;
	Matches 6; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
DY	2 SERVDR 7	
Db	545 sfrvdgr 550	
RESULT 8		
AAB19802		
ID AAB19802 standard; Protein; 1576 AA.		
AC AAB19802;		
DT 05-MAR-2001 (first entry)		
DE Human laminin 2 mature gamma-1 chain.		
KM laminin 2; human; nerve regeneration; angiogenic; cell adhesion;		
KM degenerative muscle disorder; muscular dystrophy; cell therapy.		
OS Homo sapiens.		

PN WO200066730-A2.  
XX  
PD 09-NOV-2000.  
XX  
PF 28-APR-2000; 2000WO-US11378.  
XX  
PR 30-APR-1999; 99US-0131720.  
PR 15-JUN-1999; 99US-0139198.  
PR 12-JUL-1999; 99US-0143289.  
PR 24-SEP-1999; 99US-0155945.  
XX  
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
XX  
PI Yurchenco P;  
XX  
DR WPI: 2000-687537/67.  
DR N-PSDB: AAA88902.  
XX  
PT Purified laminin 2 protein, useful for research and therapeutic purposes including peripheral nerve regeneration, treatment of degenerative muscle disorders, angiogenesis regulation, and ex vivo cell therapy -  
XX  
PS Claim 5; Page 251-256; 305pp; English.  
XX  
CC The present sequence is that of human laminin 2 gamma-1 chain mature protein. laminin-2 is composed of alpha-2 (400 kDa), beta-1 (100 kDa) and gamma-1 (100 kDa) chains. It is thought to be specifically required for stabilizing myotubes during skeletal muscle development, and for preventing apoptosis. Genetic defects in its structure or expression are associated with a major type of congenital muscular dystrophy. Laminin 2 is also thought to be important in Schwann cell/basal lamina interactions. The invention provides laminin 2 alpha-2, beta-1 and gamma-1 chain polypeptides (see AAB19791-806) and the polynucleotides encoding them (see AAA8891-906), methods for making recombinant laminin 2, cells that express recombinant laminin 2, and methods for using purified laminin 2 for research and therapeutic purposes including peripheral nerve regeneration, treatment of degenerative muscle disorders, angiogenesis regulation, promoting cell attachment and migration, ex vivo cell therapy, improving the take of grafts, improving the biocompatibility of medical devices and preparing improved culture devices and media.  
CC  
XX  
SQ Sequence 1576 AA;  
XX  
Query Match 88.2%; Score 30; DB 21; Length 1576;  
Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SERVDR 7  
DB 547 sftvdr 552  
XX  
RESULT 9  
AAB48453  
ID AAB48453 standard; Protein: 1576 AA.  
XX  
AC AAB48453;  
XX  
DT 02-MAR-2001 (first entry)  
XX  
DE Human laminin 8 polypeptide; SEQ ID NO: 24.  
XX  
KW Human; laminin 8; neuroprotective; angiogenic; osteopathic; antiatherosclerotic; glycoprotein; mesenchymal tissue injury; vascular tissue injury; neural injury; angiogenesis regulation.  
XX  
OS Homo sapiens.  
XX  
PN WO200066732-A2.

XX  
PD 09-NOV-2000.  
XX  
PF 28-APR-2000; 2000WO-US11543.  
XX  
PR 30-APR-1999; 99US-0131720.  
PR 21-AUG-1999; 99US-0149738.  
PR 24-SEP-1999; 99US-0155945.  
PR 11-FEB-2000; 2000US-0182012.  
XX  
PA (BIOS-) BIOSTRATUM INC.  
XX  
PI Kortessmaa J, Tryggvason K;  
XX  
DR WPI: 2000-687539/67.  
DR N-PSDB: AAC83714.  
XX  
PT Purified laminin 8 protein, useful for research and therapeutic purposes including peripheral nerve regeneration, treatment of degenerative muscle disorders, angiogenesis regulation, and ex vivo cell therapy -  
XX  
PS Claim 5; Page 214-218; 245pp; English.  
XX  
CC The present sequence is a laminin 8 polypeptide chain. Laminins are a family of heterotrimeric glycoproteins that function via binding interactions with neighbouring cell receptors and by forming laminin networks. They are signalling molecules which influence cellular function. Laminin 8 is useful for treating injuries to tissue of mesenchymal origin, such as bone, cartilage, tendon, and ligament, treating injuries to vascular tissue, promoting cell attachment and migration, ex vivo cell therapy, improving the biocompatibility of medical devices, and preparing improved cell culture devices and media. Laminin 8 is also useful for promoting re-neurocellularisation at the site of vascular injuries, improving the take of grafts, improving the biocompatibility of medical devices, treating neural injuries (neural regeneration), regulating angiogenesis, and promoting cell attachment and migration.  
CC  
XX  
SQ Sequence 1576 AA;  
XX  
Query Match 88.2%; Score 30; DB 21; Length 1576;  
Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SERVDR 7  
DB 547 sftvdr 552  
XX  
RESULT 10  
AAB19804  
ID AAB19804 standard; Protein: 1584 AA.  
XX  
AC AAB19804;  
XX  
DT 05-MAR-2001 (first entry)  
XX  
DE Human laminin 2 gamma-1 chain with C-terminal FLAG epitope.  
XX  
KW Laminin 2; human; nerve regeneration; angiogenic; cell adhesion; degenerative muscle disorder; muscular dystrophy; cell therapy.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FH Peptide 1..33 Location/Qualifiers  
FT Protein /label= Signal\_peptide 34..1609  
FT Peptide /label= Mature\_protein 1610..1617  
FT /label= FLAG

```
XX XX WO200066730-A2.
PN XX
XX XX
PD PD 09-NOV-2000.
XX XX
PF 28-APR-2000; 2000MO-US11378.
XX XX
PR 30-APR-1999; 990S-0131720.
PR 15-JUN-1999; 990S-0139198.
PR 12-JUL-1999; 990S-0143289.
PR 24-SEP-1999; 990S-0155945.
XX XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX XX Yurchenco P;
PI
XX XX WPI: 2000-687537/67.
DR N-PSDB; AAA88904.
XX XX
XX XX Purified laminin 2 protein, useful for research and therapeutic
PT purposes including peripheral nerve regeneration, treatment of
PT degenerative muscle disorders, angiogenesis regulation, and ex vivo
XX cell therapy -
XX
PS Claim 5; Page 275-280; 305pp; English.
XX
XX The present sequence is that of the mature gamma-1 chain of human
CC laminin 2, with an additional C-terminal FLAG epitope, resulting from
CC expression in transfected cells from mammalian expression vectors.
CC Laminin 2 is composed of alpha-2 (400 kDa), beta-1 (100 kDa) and
CC gamma-1 (100 kDa) chains. It is thought to be specifically
CC required for stabilizing myotubes during skeletal muscle
CC development, and for preventing apoptosis. Genetic defects in its
CC structure or expression are associated with a major type of
CC congenital muscular dystrophy. Laminin 2 is also thought to be
CC important in Schwann cell/basal lamina interactions. The invention
CC provides laminin 2 alpha-2, beta-1 and gamma-1 chain polypeptides
CC (see AAB19791-806) and the polynucleotides encoding them (see
CC AAA88891-906), methods for making recombinant laminin 2, cells that
CC express recombinant laminin 2, and methods for using purified
CC laminin 2 for research and therapeutic purposes including peripheral
CC nerve regeneration, treatment of degenerative muscle disorders,
CC angiogenesis regulation, promoting cell attachment and migration,
CC ex vivo cell therapy, improving the take of grafts, improving the
CC biocompatibility of medical devices and preparing improved culture
CC devices and media.
XX
SO Sequence 1584 AA;

Query Match 88.2%; Score 30; DB 21; Length 1584;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7
DB 547 stvdr 552

RESULT 11
AAB19805
ID AAB19805 standard; Protein; 1605 AA.
XX
XX
AC AAB19805;
XX
XX
DT 05-MAR-2001 (first entry)
XX
XX Mouse laminin 2 gamma-1 chain.
XX
XX Laminin 2; mouse; nerve regeneration; angiogenic; cell adhesion;
KM degenerative muscle disorder; muscular dystrophy; cell therapy.
XX
OS Mus musculus.
```

```
XX XX
FH Key Location/Qualifiers
FT Peptide 1..33
FT /label= Signal_peptide
FT Protein 34..1605
FT /label= Mature_protein
XX
XX WO200066730-A2.
PN XX
XX XX
PD PD 09-NOV-2000.
XX XX
PF 28-APR-2000; 2000MO-US11378.
XX XX
PR 30-APR-1999; 990S-0131720.
PR 15-JUN-1999; 990S-0139198.
PR 12-JUL-1999; 990S-0143289.
PR 24-SEP-1999; 990S-0155945.
XX XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX XX Yurchenco P;
PI
XX XX WPI: 2000-687537/67.
DR N-PSDB; AAA88905.
XX XX
XX XX Purified laminin 2 protein, useful for research and therapeutic
PT purposes including peripheral nerve regeneration, treatment of
PT degenerative muscle disorders, angiogenesis regulation, and ex vivo
XX cell therapy -
XX
PS Claim 5; Page 288-294; 305pp; English.
XX
XX The present sequence is that of the gamma-1 chain of mouse laminin 2.
CC Laminin 2 is composed of alpha-2 (400 kDa), beta-1 (100 kDa) and
CC gamma-1 (100 kDa) chains. It is thought to be specifically
CC required for stabilizing myotubes during skeletal muscle
CC development, and for preventing apoptosis. Genetic defects in human
CC laminin 2 structure or expression are associated with a major type
CC of congenital muscular dystrophy. Laminin 2 is also thought to be
CC important in Schwann cell/basal lamina interactions. The invention
CC provides laminin 2 alpha-2, beta-1 and gamma-1 chain polypeptides
CC (see AAB19791-806) and the polynucleotides encoding them (see
CC AAA88891-906), methods for making recombinant laminin 2, cells that
CC express recombinant laminin 2, and methods for using purified
CC laminin 2 for research and therapeutic purposes including peripheral
CC nerve regeneration, treatment of degenerative muscle disorders,
CC angiogenesis regulation, promoting cell attachment and migration,
CC ex vivo cell therapy, improving the take of grafts, improving the
CC biocompatibility of medical devices and preparing improved culture
CC devices and media.
XX
SO Sequence 1605 AA;

Query Match 88.2%; Score 30; DB 21; Length 1605;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7
DB 578 stvdr 583

RESULT 12
AAB48454
ID AAB48454 standard; Protein; 1605 AA.
XX
XX
AC AAB48454;
XX
XX
DT 02-MAR-2001 (first entry)
XX
XX Mouse laminin 8 polypeptide, SEQ ID NO: 26.
XX
```

KW	Carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
KM	Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
KW	therapy.
XX	
OS	Mus sp.
PN	MO9815179-A1.
PD	
XX	16-APR-1998.
PF	
XX	08-OCT-1997; 97WO-US18145.
XX	
PR	08-OCT-1996; 96US-0027981.
XX	
PA	(UNIW ) UNIV WASHINGTON.
XX	
PI	Castillo G, Snow AD:
DR	WPI: 1998-240534/21.
PT	
PT	Use of laminin and fragments - for developing products for use in the diagnosis and treatment of amyloid disease, e.g. Alzheimer's disease or CJD
PS	
XX	Claim 15; Page 102-105; 132pp; English.
CC	
CC	This is the amino acid sequence of the mouse laminin G1 chain. The
CC	primary object of the invention is to use laminin, laminin-derived
CC	protein fragments and/or laminin-derived polypeptides as potent
CC	inhibitors of amyloid formation, deposition, accumulation and/or
CC	persistence in Alzheimer's disease and other amyloidoses. The
CC	laminin products (see AAW50888-98) may include mouse or human laminin
CC	A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merostin),
CC	laminin G1 chain, the globular repeats of the laminin A1 chain and
CC	the beta-amyloid binding domain of the laminin A chain. A claimed
CC	method for treating an amyloid disease comprises administering a
CC	polypeptide having a conformational similarity to a fragment of a
CC	laminin protein. A method for diagnosing an amyloid disease
CC	involves determining levels of laminin in a sample. Production
CC	of laminin or its fourth globular repeat in vivo provides a method
CC	for in vivo inhibition of beta-amyloid amyloidosis. The products
CC	and methods can be used for the diagnosis, prognosis, monitoring
CC	and treatment of amyloidoses such as Alzheimer's disease, Down's
CC	syndrome and hereditary cerebral haemorrhage with amyloidosis of
CC	the Dutch type (where the specific amyloid is the beta-amyloid
CC	protein), the amyloidosis associated with chronic inflammation,
CC	various forms of malignancy and Familial Mediterranean Fever (AA
CC	amyloid or inflammation-associated amyloidosis), the amyloidosis
CC	associated with multiple myeloma and other B-cell abnormalities
CC	(AL amyloid), the amyloidosis associated with type II diabetes
CC	(amylin or islet amyloid), the amyloidosis associated with prion
CC	diseases including Creutzfeldt-Jacob disease, Gerstmann-Straussler
CC	syndrome, kuru and animal scrapie (PP amyloid), the amyloidosis
CC	associated with long-term haemodialysis and carpal tunnel syndrome
CC	(beta-2-microglobulin amyloid), the amyloidosis associated with
CC	senile cardiac amyloid and Familial Amyloidotic Polynuropathy
CC	(prealbumin or transthyretin amyloid), and the amyloidosis
CC	associated with endocrine tumours such as medullary carcinoma of
CC	the thyroid (variant of procalcitonin).
XX	
SQ	Sequence 1607 AA:
Query Match	88.2%, Score 30; DB 19; Length 1607;
Best Local Similarity	100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	2 SERVDR 7
Db	 578 SERVDR 583

AAW50898  
ID AAW50898 standard; Protein; 1609 AA.  
XX  
AC AAW50898;  
XX  
DT 07-DEC-1998 (first entry)  
XX  
DE Human laminin G1 chain.  
XX  
KW Laminin; human; beta-amyloid; amyloidosis; Alzheimer's disease;  
KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;  
KW malignancy; Familial Mediterranean Fever; multiple myeloma;  
KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;  
KW Gertsman-Straussler syndrome; kuru; scrapie; haemodialysis;  
KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;  
KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;  
KW therapy.  
XX  
OS Homo sapiens.  
XX  
PN W09815179-A1.  
XX  
PD 16-APR-1998.  
XX  
PE 08-OCT-1997; 97WO-US18145.  
XX  
PR 08-OCT-1996; 96US-0027981.  
XX  
PA (UNIM ) UNIV WASHINGTON.  
XX  
PI Castillo G, Snow AD;  
XX  
DR WPI; 1998-240534/21.  
XX  
PT Use of laminin and fragments - for developing products for use in  
PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's  
PT disease or CJD  
PS  
PS Claim 15; Page 106-109; 132pp; English.  
XX  
XX This is the amino acid sequence of the human laminin G1 chain. The  
CC primary object of the invention is to use laminin, laminin-derived  
CC protein fragments and/or laminin-derived polypeptides as potent  
CC inhibitors of amyloid formation, deposition, accumulation and/or  
CC persistence in Alzheimer's disease and other amyloidoses. The  
CC laminin products (see AAW50888-98) may include mouse or human laminin  
CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (nerosin),  
CC laminin G1 chain, the globular repeats of the laminin A1 chain and  
CC the beta-amyloid binding domain of the laminin A chain. A claimed  
CC method for treating an amyloid disease comprises administering a  
CC polypeptide having a conformational similarity to a fragment of a  
CC laminin protein. A method for diagnosing an amyloid disease  
CC involves determining levels of laminin in a sample. Production  
CC of laminin or its fourth globular repeat in vivo provides a method  
CC for in vivo inhibition of beta-amyloid amyloidosis. The products  
CC and methods can be used for the diagnosis, prognosis, monitoring  
CC and treatment of amyloidoses such as Alzheimer's disease, Down's  
CC syndrome and hereditary cerebral haemorrhage with amyloidosis of  
CC the Dutch type (where the specific amyloid is the beta-amyloid  
CC protein), the amyloidosis associated with chronic inflammation,  
CC various forms of malignancy and Familial Mediterranean Fever (AA  
CC amyloid or inflammation-association amyloidosis), the amyloidosis  
CC associated with multiple myeloma and other B-cell abnormalities  
CC (AL amyloid), the amyloidosis associated with type II diabetes  
CC (amylin or islet amyloid), the amyloidosis associated with prion  
CC diseases including Creutzfeldt-Jacob disease, Gertsman-Straussler  
CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis  
CC associated with long-term haemodialysis and carpal tunnel syndrome  
CC (beta 2-microglobulin amyloid), the amyloidosis associated with  
CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy  
CC (prealbumin or transthyretin amyloid), and the amyloidosis  
CC associated with endocrine tumours such as medullary carcinoma of  
CC the thyroid (variant of procalcitonin).  
4

XX  
SQ Sequence 1609 AA;  
XX  
Query Match 88.2%; Score 30; DB 19; Length 1609;  
Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SFRVDR 7  
Db 580 sfrvdr 585  
RESULT 15  
AAB19801  
ID AAB19801 standard; Protein; 1609 AA.  
XX  
AC AAB19801;  
XX  
DT 05-MAR-2001 (first entry)  
XX  
DE Human laminin 2 gamma-1 chain.  
XX  
KW Laminin 2; human; nerve regeneration; angiogenic; cell adhesion;  
KW degenerative muscle disorder; muscular dystrophy; cell therapy.  
XX  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..33  
FT /label= Signal\_peptide  
FT Protein 34..1609  
FT /label= Mature\_protein  
XX  
PN W0200066730-A2.  
XX  
PD 09-NOV-2000.  
XX  
PF 28-APR-2000; 2000WO-US11378.  
XX  
PR 30-APR-1999; 99US-0131720.  
PR 15-JUN-1999; 99US-0139198.  
PR 12-JUL-1999; 99US-0143289.  
PR 24-SEP-1999; 99US-0155945.  
XX  
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
XX  
PI Yurchenco P;  
XX  
DR WPI; 2000-687537/67.  
DR N-PSDB; AAB88901.  
XX  
PT Purified laminin 2 protein, useful for research and therapeutic  
PT purposes including peripheral nerve regeneration, treatment of  
PT degenerative muscle disorders, angiogenesis regulation, and ex vivo  
PT cell therapy -  
PS  
PS Claim 5; Page 239-244; 305pp; English.  
XX  
XX The present sequence is that of the gamma-1 chain of human laminin 2.  
CC Laminin 2 is composed of alpha-2 (400 kDa), beta-1 (100 kDa) and  
CC gamma-1 (100 kDa) chains. It is thought to be specifically  
CC required for stabilizing myotubes during skeletal muscle  
CC development, and for preventing apoptosis. Genetic defects in its  
CC structure or expression are associated with a major type of  
CC congenital muscular dystrophy. Laminin 2 is also thought to be  
CC important in Schwann cell/basal lamina interactions. The invention  
CC provides laminin 2 alpha-2, beta-1 and gamma-1 chain polypeptides  
CC (see AAB19791-806) and the polynucleotides encoding them (see  
CC AAB8891-906), methods for making recombinant laminin 2, cells that  
CC express recombinant laminin 2, and methods for using purified  
CC laminin 2 for research and therapeutic purposes including peripheral  
CC nerve regeneration, treatment of degenerative muscle disorders,



CC angiogenesis regulation, promoting cell attachment and migration,  
CC ex vivo cell therapy, improving the take of grafts, improving the  
CC biocompatibility of medical devices and preparing improved culture  
CC devices and media.  
XX

SQ Sequence 1609 AA;

Query Match 88.2%; Score 30; DB 21; Length 1609;

Best Local Similarity 100.0%; Pred. NO. 5.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFRVDR 7

Db 580 sfrvdr 585

Search completed: May 24, 2002, 16:48:08  
Job time: 248 sec

THIS PAGE BLANK (USTP)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:10 ; Search time 107.9 Seconds  
(without alignments)  
120.247 Million cell updates/sec

Title: US-09-730-379e-1  
Perfect score: 424  
Sequence: 1 GPRFHCROIGSVYRLPLR.....SESCPGKKGFPQVSMFF 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	271	63.9	525	11	Q99PS6	Q99PS6 mus musculus
2	263	62.0	515	11	Q99PS7	Q99PS7 ratius norv
3	263	62.0	525	11	Q99PS3	Q99PS3 mus musculus
4	263	62.0	525	11	Q99PS5	Q99PS5 mus musculus
5	249	58.7	510	11	Q99SB2	Q99SB2 ratius norv
6	249	58.7	525	11	Q99PS8	Q99PS8 ratius norv
7	240	56.6	546	6	Q99G01	Q99G01 bos taurus
8	77.5	18.3	260	10	Q40053	Q40053 hordeum vul
9	76.5	18.0	313	10	Q39333	Q39333 brassica na
10	76	17.9	279	10	Q41543	Q41543 triticum ae
11	76	17.9	298	10	Q94G92	Q94G92 triticum ae
12	75	17.7	310	10	Q41210	Q41210 hordeum vul
13	75	17.7	456	2	Q92BP2	Q92BP2 streptomyce
14	74.5	17.6	740	3	Q96U76	Q96U76 neurospora
15	74	17.5	241	10	Q94G98	Q94G98 triticum ae
16	74	17.5	337	10	Q94G96	Q94G96 triticum ae

17	74	17.5	446	5	Q9VZC2	Q9VZC2 drosophila
18	74	17.5	530	5	Q9YQ34	Q9YQ34 drosophila
19	74	17.5	536	5	Q961I7	Q961I7 drosophila
20	73.5	17.3	357	10	Q43639	Q43639 secale cere
21	73.5	17.3	412	11	Q70450	Q70450 mus musculus
22	73.5	17.3	538	5	Q24615	Q24615 drosophila
23	73.5	17.3	557	11	Q921J1	Q921J1 mus musculus
24	73.5	17.3	584	11	Q70573	Q70573 mus musculus
25	73	17.2	192	10	Q9FS58	Q9FS58 triticum ur
26	73	17.2	255	10	Q9FTC2	Q9FTC2 aegilops sp
27	73	17.2	347	10	Q40055	Q40055 hordeum vul
28	72	17.0	201	10	Q9FTB4	Q9FTB4 aegilops ta
29	72	17.0	203	10	Q9FTB5	Q9FTB5 aegilops ta
30	72	17.0	204	10	Q9FTB7	Q9FTB7 aegilops ta
31	72	17.0	207	10	Q9FTB2	Q9FTB2 aegilops ta
32	72	17.0	234	10	Q9FTC3	Q9FTC3 aegilops sh
33	72	17.0	243	10	Q9FTC0	Q9FTC0 aegilops ta
34	72	17.0	243	10	Q9FTB9	Q9FTB9 aegilops ta
35	72	17.0	244	10	Q9FS61	Q9FS61 triticum ae
36	72	17.0	248	10	Q9FS57	Q9FS57 triticum va
37	72	17.0	250	10	Q9FS76	Q9FS76 triticum ae
38	72	17.0	252	10	Q9FS77	Q9FS77 triticum ae
39	72	17.0	254	10	Q9FS56	Q9FS56 triticum va
40	72	17.0	259	10	Q9FS75	Q9FS75 triticum ae
41	72	17.0	259	10	Q9FS62	Q9FS62 triticum ae
42	72	17.0	264	10	Q9FS74	Q9FS74 triticum ae
43	72	17.0	267	10	Q9FBS9	Q9FBS9 triticum ae
44	72	17.0	267	10	Q9FS60	Q9FS60 triticum ae
45	72	17.0	270	10	Q9FS69	Q9FS69 triticum ae

#### ALIGNMENTS

RESULT 1  
Q99PS6 PRELIMINARY: PRT: 525 AA.  
ID Q99PS6:  
AC Q99PS6:  
DT 01-JUN-2001 (TREMBL) 17, Created  
DT 01-JUN-2001 (TREMBL) 17, Last sequence update  
DT 01-OCT-2001 (TREMBL) 18, Last annotation update  
DE HISTIDINE-RICH GLYCOPROTEIN.  
GN MHRG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE=LIVER;  
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
RA Onashi T., Sato N., Hirata D., Tsuchida N., Koide F.,  
RT "Molecular diversity of mammalian histidine-rich glycoprotein."  
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AB058897; BAB33094.1; -  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; Cystatin; 1.  
DR SMART; SM00043; CT; 2.  
SQ SEQUENCE 525 AA; 59090 MW; A83E93A439CFB3AC CRC64;

Query Match 63.9%; Score 271; DB 11; Length 525;  
Best Local Similarity 70.4%; Pred. No. 1.5e-24;  
Matches 50; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 4 PFCRCROIGSVYRLPLRKGEVLPLRPNKPLKPDNPFGQVSESCPGKF 63  
DB 449 PFCRCROIGSVYRLPLRKGEVLPLRPNKPLKPDNPFGQVSESCPGKF 508  
QY 64 KSGFPQVSMFF 74  
DB 509 ESGFPQVSMFF 519

RESULT 2  
 ID Q99PS7 PRELIMINARY; PRT; 515 AA.  
 AC Q99PS7;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN 2.  
 GN RNRHG2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB055896; BAB33093.1; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; C1; 2.  
 SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 62.0%; Score 263; DB 11; Length 515;  
 Best Local Similarity 70.4%; Pred. NO.1.4e-23;  
 Matches 50; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 4 PFHCROIGSVYRLPPLRKKEVLTLPPEANFPSPFLPHHKHPLKPDNPQPPQSVSSSCPGKF 63  
 Db 439 PFHQROIGYVYRLPPLNIGEVLTLPPEANFPSPFLPNCNRPQPIQGFPPQASNSCPGKF 498  
 QY 64 KSGFPQVSMFP 74  
 Db 499 ESKFPQVSNFP 509

RESULT 3  
 ID Q9ESB3 PRELIMINARY; PRT; 525 AA.  
 AC Q9ESB3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN.  
 GN HRG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=20307726; PubMed=10849117;  
 RA Hulet M.D., Parish C.R.;  
 RT "Murine histidine-rich glycoprotein: cloning, characterization and  
 RT cellular origin.";  
 RL Immunol. Cell Biol. 78:280-287(2000).  
 DR EMBL; AF194028; AAG28416.1; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; C1; 2.  
 SQ SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;

Query Match 62.0%; Score 263; DB 11; Length 525;  
 Best Local Similarity 69.0%; Pred. NO.1.4e-23;  
 Matches 49; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 4 PFHCROIGSVYRLPPLRKKEVLTLPPEANFPSPFLPHHKHPLKPDNPQPPQSVSSSCPGKF 63  
 Db 449 PFHQROIGYVYRLPPLNIGEVLTLPPEANFPSPFLPNCNRPQPIQGFPPQASNSCPGKF 508  
 QY 64 KSGFPQVSMFP 74  
 Db 509 ESKFPQVSNFP 519

RESULT 4  
 ID Q99PS5 PRELIMINARY; PRT; 525 AA.  
 AC Q99PS5;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).  
 GN MHRG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tsuchida N., Wakabayashi S., Jahnke-Dechent W., Koide T.;  
 RT "Structure of mouse histidine-rich glycoprotein gene.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB055898; BAB33095.1; -;  
 DR EMBL; BC011168; AAH11168.1; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; C1; 2.  
 SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match 62.0%; Score 263; DB 11; Length 525;  
 Best Local Similarity 69.0%; Pred. NO.1.4e-23;  
 Matches 49; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 4 PFHCROIGSVYRLPPLRKKEVLTLPPEANFPSPFLPHHKHPLKPDNPQPPQSVSSSCPGKF 63  
 Db 449 PFHQROIGYVYRLPPLNIGEVLTLPPEANFPSPFLPNCNRPQPIQGFPPQASNSCPGKF 508  
 QY 64 KSGFPQVSMFP 74  
 Db 509 ESKFPQVSNFP 519  
 RESULT 5  
 ID Q9ESB2 PRELIMINARY; PRT; 510 AA.  
 AC Q9ESB2;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN.  
 GN HRG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LEWIS;  
 RX MEDLINE=20307726; PubMed=10849117;  
 RA Hulet M.D., Parish C.R.;  
 RT "Murine histidine-rich glycoprotein: cloning, characterization and  
 RT cellular origin.";

RL Immunol. Cell Biol. 78:280-287(2000).  
 DR EMBL: AF194029; AAG28417.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; Cy; 2.  
 SQ SEQUENCE 510 AA; 57581 MW; 508E606AA2ED58E CRC64;

Query Match 58.7%; Score 249; DB 11; Length 510;  
 Best Local Similarity 66.2%; Pred. No. 6.6e-22;  
 Matches 47; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 4 PFHCRIGSVYRLPLRKGEVLPLPEANFSPPLPHKHKPLKPDNPQSPQSVSESCPGKF 63  
 DB 434 PFHCRIGSVYRLPLPLNGEVLPLPEANFIFSLPCNCRPQPEIRPFQATSKSCPGKF 493  
 QY 64 KSGFPQVSMEF 74  
 DB 494 ECKFPQVSMEF 504

RESULT 6  
 Q99PS8 PRELIMINARY: PRT; 525 AA.

AC Q99PS8;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN 1.  
 GN RNHRG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.,  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein."  
 RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB055895; BAB33092.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; Cy; 2.  
 SQ SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match 58.7%; Score 249; DB 11; Length 525;  
 Best Local Similarity 66.2%; Pred. No. 6.9e-22;  
 Matches 47; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 4 PFHCRIGSVYRLPLRKGEVLPLPEANFSPPLPHKHKPLKPDNPQSPQSVSESCPGKF 63  
 DB 449 PFHCRIGSVYRLPLPLNGEVLPLPEANFIFSLPCNCRPQPEIRPFQATSKSCPGKF 508  
 QY 64 KSGFPQVSMEF 74  
 DB 509 ECKFPQVSMEF 519

RESULT 7  
 Q9BGU1 PRELIMINARY: PRT; 546 AA.  
 AC Q9BGU1;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE HISTIDINE-RICH GLYCOPROTEIN.  
 GN BTHRG.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

OC Bovidae; Bovinae; Bos.  
 RN NCBI\_TaxID=9913;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
 RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.,  
 RT "Molecular diversity of mammalian histidine-rich glycoprotein."  
 RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB055894; BAB33091.1; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; Cy; 2.  
 SQ SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 56.6%; Score 240; DB 6; Length 546;  
 Best Local Similarity 62.9%; Pred. No. 8.8e-21;  
 Matches 44; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

QY 5 PFHCRIGSVYRLPLRKGEVLPLPEANFSPPLPHKHKPLKPDNPQSPQSVSESCPGKF 64  
 DB 471 FFWRPYGYIHRPLSLKKEVLPLPEANFSPSLPHNNPLQPIQAFPOSASESCPGTN 530  
 QY 65 SGEFPQVSMEF 74  
 DB 531 IKFLHISKEF 540

RESULT 8  
 Q40053 PRELIMINARY: PRT; 260 AA.

AC Q40053;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HOR-17 C-HORDEIN.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticaceae; Hordeum.  
 RN NCBI\_TaxID=4513;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BOMI;  
 RX MEDLINE=92032786; PubMed=1932695;  
 RA Entwistle J., Knudsen S., Mueller M., Cameron-Mills V.,  
 RT "Amber codon suppression: the in vivo and in vitro analysis of two C-  
 RT hordein genes from barley."  
 RT Plant Mol. Biol. 17:1217-1231(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BOMI;  
 RX MEDLINE=94035184; PubMed=8220485;  
 RA Mueller M., Knudsen S.,  
 RT "The nitrogen response of a barley C-hordein promoter is controlled by  
 RT positive and negative regulation of the GCN4 and endosperm box."  
 RT Plant J. 4:343-355(1993).  
 DR EMBL: X60037; CAA42642.1; -  
 SQ SEQUENCE 260 AA; 30397 MW; CB220BD2309280AE CRC64;

Query Match 18.3%; Score 77.5; DB 10; Length 260;  
 Best Local Similarity 27.7%; Pred. No. 0.17;  
 Matches 26; Conservative 11; Mismatches 28; Indels 29; Gaps 5;

QY 3 RPFHCRQ-----IGSVYRLP-----PLRKGEVLPLPEANF-----SFLPH 39  
 DB 57 QPHTHQYQYFYLPELFPQYQIPLQPOPPQPPQPPQPPQPPQPPQPPQPPQPPQ 116  
 QY 40 HKHPLKPDNPQSPQSVSESCPGKFKSGFPQVSMEF 73  
 DB 117 EP1POOP-QQPPQPOPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 144

```

RESULT 9
Q39333 PRELIMINARY; PRT; 313 AA.
ID Q39333
AC Q39333
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE G-BOX BINDING FACTOR 1A.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
ON NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COTYLEDONS;
RX MEDLINE=96123230; PubMed=8534849;
RA Waldmüller S., Link G.;
RT "Sequence and expression characteristics of three g-Box-binding factor
RT cdnas from Brassica napus."
RL Plant Mol. Biol. 29:507-517(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL: X83922; CA58774.1; -.
DR HSSP: P03069; IYSA.
DR TRANSFAC: T02798; -.
DR InterPro: IPR001871; bZIP.
DR Pfam: PF00170; bZIP; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 313 AA; 33645 MW; F431C419FA9E1CEF CRC64;

Query Match 18.0%; Score 76.5; DB 10; Length 313;
Best Local Similarity 27.6%; Pred. No. 0.27; Mismatches 34; Indels 13; Gaps 2;
Matches 21; Conservative 8;

QY 2 PRPFHCQIGSV-----YRLPRLKGEVLPRLPEANFSPRLPHKHPKLPDNP 50
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 41 PSFFPSPVSPSPHRYMGAGNHMMRPY--GPRVPRPAMYPRGAVYAHVMPWPSSAP 98
QY 51 FPOSVSSESCGPKFKSG 66
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 99 TNETVKEQAPGKKSKG 114

RESULT 10
Q41543 PRELIMINARY; PRT; 279 AA.
ID Q41543
AC Q41543
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE GAMMA-GLIADIN (FRAGMENT).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triliceae; Triticum.
ON NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. HARD RED SPRING 10W;
RX MEDLINE=98409296; PubMed=9738916;
RA Maruyama N., Ichise K., Katsube T., Kishimoto T., Kawase S.,
RA Matsumura Y., Takeuchi Y., Sawada T., Utsumi S.;
RT "Identification of major wheat allergens by means of the Escherichia
RT coli expression system."
RL Eur. J. Biochem. 255:739-745(1998).
CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOG
  CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO SINGLE-

```

```

CC BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR
CC DIRECT REPEATS.
DR EMBL: D78183; BAAL1251.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amy1_inh.
DR InterPro: IPR001954; Glia-glutenin.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00208; GLIADGLUTEN.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Repeat; Multigene family.
FT NON_TER
SQ SEQUENCE 279 AA; 31852 MW; 83CCRF1E52020B02 CRC64;

Query Match 17.9%; Score 76; DB 10; Length 279;
Best Local Similarity 30.2%; Pred. No. 0.28; Mismatches 24; Indels 6; Gaps 1;
Matches 16; Conservative 7;

QY 23 EYLRLPEANFSPRLPHKHPKLPD-----NDFPOSVSSESCGPKFKSGFPQ 69
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 31 QTFPPOQOTFRPHQPOQFOFQPOQFOQFIOPOQFPFOQPOQOTYPORPQOFPPQ 83

RESULT 11
Q94G92 PRELIMINARY; PRT; 298 AA.
ID Q94G92
AC Q94G92
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE GAMMA-GLIADIN.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triliceae; Triticum.
ON NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RA Anderson O.D., Hsia C., Torres V.;
RT "The wheat gamma-gliadin genes: characterization of ten new sequences
RT and further understanding of gamma-gliadin gene family structure."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF234649; AAK84778.1; -.
SQ SEQUENCE 298 AA; 33893 MW; DC3E5803CF4C2724 CRC64;

Query Match 17.9%; Score 76; DB 10; Length 298;
Best Local Similarity 30.2%; Pred. No. 0.3; Mismatches 24; Indels 6; Gaps 1;
Matches 16; Conservative 7;

QY 23 EYLRLPEANFSPRLPHKHPKLPD-----NDFPOSVSSESCGPKFKSGFPQ 69
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 50 QTFPPOQOTFRPHQPOQFOFQPOQFOQFIOPOQFPFOQPOQOTYPORPQOFPPQ 102

RESULT 12
Q41210 PRELIMINARY; PRT; 310 AA.
ID Q41210
AC Q41210
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE C-HORDEIN.
GN C-HORDEIN.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triliceae; Hordeum.
ON NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93380629; PubMed=8396543;

```

RA Sainova O.V., Mekhedov S.L., Zhelnin L.G., Khokhlova T.A.,  
RT Anan'ev E.V.;  
RL "[Nucleotide sequence of the barley C-hordein gene].";  
DR EMBL:566938; AAB28161.1; -;  
SQ SEQUENCE 310 AA; 36508 MW; 7A9621869D50D672 CRC64;

Query Match 17.7%; Score 75; DB 10; Length 310;  
Best Local Similarity 26.7%; Pred. No. 0.41; Mismatches 31; Indels 24; Gaps 4;  
Matches 24; Conservative 11; Mismatches 31; Indels 24; Gaps 4;

QY 3 REPFCRQ-----IGSVYRLP-----PLRKGEVLPLEANFP-----SFLPH 39  
DB 57 QPFTHPQYFPLPELSQYQIFPLPQRPQRPQRPQRPQRPQRPQRPQRPQ 116  
QY 40 HNHPLKPDNPQSVSESCPGKFGSPQ 69  
DB 117 QPFPYQP-QQPFNQPOQIISQOQPQFPQ 145

RESULT 13  
Q92BP2 PRELIMINARY; PRT; 456 AA.  
AC Q92BP2;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE 50K PROLINE RICH PROTEIN.  
GN SC6C5.03.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA MEDLINE-97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RL "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
DR EMBL:AL034492; CA22501.1; -;  
DR InterPro: IPR002965; P\_Rich\_extensn.  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 456 AA; 50032 MW; 1293FA497256AC57 CRC64;

Query Match 17.7%; Score 75; DB 2; Length 456;  
Best Local Similarity 33.3%; Pred. No. 0.63;  
Matches 17; Conservative 8; Mismatches 14; Indels 12; Gaps 2;

QY 2 PRPFHCRQGSYVRLPLRKGEVLPLEANFPSPFLPHKHPKLPKPDNPQFP 52  
DB 199 PRPYG-----RRPFR-----PYPVSHVYTRPPTSDPIRPNRPV 237

RESULT 14  
Q96U76 PRELIMINARY; PRT; 740 AA.  
AC Q96U76;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 81.7 KDA PROTEIN.  
GN B18D24.090.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL:AL514466; CAD11415.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 740 AA; 81656 MW; 2DB631C9247E890B CRC64;

Query Match 17.6%; Score 74.5; DB 3; Length 740;  
Best Local Similarity 26.4%; Pred. No. 1.2;  
Matches 24; Conservative 11; Mismatches 29; Indels 27; Gaps 3;

QY 2 PRPFHCRQGSYVRLPLRKGEV-----LPLPEANFPSPFL----- 37  
DB 527 PRPFHCRQGSYVRLPLRKGEV-----LPLPEANFPSPFL----- 37  
QY 38 PHKHPLKPDNPQSVSESCPGKFGSP 68  
DB 587 PHKHPLKPDNPQSVSESCPGKFGSP 68

RESULT 15  
Q94G98 PRELIMINARY; PRT; 241 AA.  
AC Q94G98;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE GAMMA-GLIADIN (FRAGMENT).  
OC Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticaceae; Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. CHEYENNE;  
RA Anderson O.D., Hsia C., Torres V.;  
RL "The wheat gamma-gliadin genes: characterization of ten new sequences  
and further understanding of gamma-gliadin gene family structure.";  
DR EMBL:AF234642; AAK84772.1; -;  
FT NON-TER 241  
SQ SEQUENCE 241 AA; 28079 MW; A2765A48A35FD6D3 CRC64;

Query Match 17.5%; Score 74; DB 10; Length 241;  
Best Local Similarity 32.1%; Pred. No. 0.41;  
Matches 17; Conservative 8; Mismatches 22; Indels 6; Gaps 2;

QY 23 EYVLPLEANFP-----SFLPHKHPKLPKPDNPQSVSESCPGKFGSPQ 69  
DB 51 QIFPQPOQTFPHQPOQAFPOQOQTFPHQPOQOQFPQOQFPQOQOQFPQ 103

Search completed: May 24, 2002, 17:02:47  
Job time: 877 sec

THIS PAGE BLANK (USPSTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:44:15 ; Search time 49.37 Seconds  
(without alignments)  
37.106 Million cell updates/sec

Title: US-09-730-379E-1

Perfect score: 424

Sequence: 1 GPRFHCROIGSYRLPLR.....SESCPGKSGFPQVSMFFT 75

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

- 1: /cgn2\_6/plodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/plodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/plodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/plodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/plodata/2/1aa/PTUG.COMB.pep:\*
- 6: /cgn2\_6/plodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	16.0	947	2	US-08-887-518-2
2	68	16.0	947	2	US-09-023-321-2
3	68	16.0	947	2	US-09-032-475-2
4	68	16.0	947	1	US-09-257-703-1
5	67	15.8	78	1	US-08-487-359-7
6	67	15.8	78	1	US-08-222-798A-7
7	65	15.3	465	4	US-08-840-767-8
8	65	15.3	465	4	US-08-840-767-8
9	64	15.1	465	4	US-08-840-767-52
10	64	15.1	465	4	US-09-096-776B-10
11	64	15.1	1105	4	US-08-999-774A-2
12	63.5	15.0	267	3	US-08-301-162-16
13	63.5	15.0	267	4	US-09-461-240-16
14	63.5	15.0	398	4	US-09-303-064-54
15	63.5	15.0	398	4	US-09-086-503-54
16	63.5	15.0	443	2	US-08-833-963C-2
17	63.5	15.0	443	3	US-08-980-514-1
18	63	14.9	590	4	US-08-893-852A-4
19	63	14.9	590	4	US-08-821-818-2
20	62.5	14.7	501	2	US-08-660-963-13
21	61.5	14.5	457	3	US-09-416-213-2
22	61	14.4	416	4	US-09-228-246-4
23	61	14.4	1824	2	US-08-680-327-3
24	61	14.4	1824	4	US-09-228-246-2
25	60	14.2	543	2	US-08-469-412A-7
26	60	14.2	543	4	US-09-021-715-7
27	60	14.2	548	2	US-08-469-412A-2

28	60	14.2	548	4	US-09-021-715-2	Sequence 2, Appl
29	60	14.2	837	3	US-09-012-710-12	Sequence 12, Appl
30	60	14.2	837	4	US-09-556-273-12	Sequence 2, Appl
31	60	14.2	1162	2	US-08-728-323A-2	Sequence 1, Appl
32	59.5	14.0	369	2	US-08-749-289-1	Sequence 1, Appl
33	59.5	14.0	657	1	US-08-264-534-34	Sequence 34, Appl
34	59.5	14.0	657	1	US-08-083-590A-13	Sequence 13, Appl
35	59.5	14.0	657	1	US-08-465-500-14	Sequence 34, Appl
36	59.5	14.0	657	2	US-08-346-128-14	Sequence 34, Appl
37	59.5	14.0	657	2	US-08-532-384-13	Sequence 13, Appl
38	59.5	14.0	657	3	US-08-893-628-14	Sequence 34, Appl
39	59.5	14.0	2471	1	US-08-185-432-16	Sequence 16, Appl
40	59.5	14.0	2471	1	US-08-083-590A-19	Sequence 19, Appl
41	59.5	14.0	2471	3	US-08-532-384-19	Sequence 19, Appl
42	59	13.9	40	3	US-08-974-775-49	Sequence 49, Appl
43	59	13.9	465	3	US-08-701-582D-13	Sequence 13, Appl
44	59	13.9	465	3	US-09-063-693A-19	Sequence 19, Appl
45	59	13.9	465	4	US-09-096-776B-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1

US-08-887-518-2

Sequence 2, Application US/08887518

Patent No. 5843721

GENERAL INFORMATION:

APPLICANT: Rothe, Mike

APPLICANT: Wu, Lin

TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/887,518

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: 797-008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341

TELEFAX: (415) 343-4342

INFORMATION FOR SEO ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 947 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-887-518-2

Query Match 16.0%; Score 68; DB 2; Length 947;

Best Local Similarity 37.7%; Pred. No. 4.5;

Matches 23; Conservative 5; Mismatches 25; Indels 8; Gaps 3;

QY 4 PFHCROIGSYRLPLRKGVLPRPEANFSPRLPH--HKHRLPLRNDQPFQSSSESCPG 61

DB 232 PLQC--LNVHVKLHHNRDGGRLPLRPTHFPPYSRLRHHPPRPRL----QPMKRPRLSESLG 285

OY 62 K 62  
DB 286 K 286

RESULT 2  
US-09-023-321-2  
; Sequence 2, Application US/09023321  
; Patent No. 5844073  
; GENERAL INFORMATION:  
; APPLICANT: Rothe, Mike  
; APPLICANT: Wu, Lin  
; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,321  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,518  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: T97-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 947 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-023-321-2

Query Match 16.0%; Score 68; DB 2; Length 947;  
Best Local Similarity 37.7%; Pred. No. 4.5;  
Matches 23; Conservative 5; Mismatches 25; Indels 8; Gaps 3;  
OY 4 PFHCROIGSVYRLPLRKEGVLPPEANPSPFLPH--HKHPLKPNQPPQSVSSCPG 61  
DB 232 PLOC--LNHWKLIHNPQDGGPLPLPTHPRPYSRLPRPFPHL-----QWKKHPLDSFLG 285  
OY 62 K 62  
DB 286 K 286  
RESULT 3  
US-09-032-475-2  
; Sequence 2, Application US/09032475  
; Patent No. 5854003  
; GENERAL INFORMATION:  
; APPLICANT: Rothe, Mike  
; APPLICANT: Wu, Lin  
; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032,475  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/887,518  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: T97-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 947 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-032-475-2

Query Match 16.0%; Score 68; DB 2; Length 947;  
Best Local Similarity 37.7%; Pred. No. 4.5;  
Matches 23; Conservative 5; Mismatches 25; Indels 8; Gaps 3;  
OY 4 PFHCROIGSVYRLPLRKEGVLPPEANPSPFLPH--HKHPLKPNQPPQSVSSCPG 61  
DB 232 PLOC--LNHWKLIHNPQDGGPLPLPTHPRPYSRLPRPFPHL-----QWKKHPLDSFLG 285  
OY 62 K 62  
DB 286 K 286

RESULT 4  
US-09-257-703-1  
; Sequence 1, Application US/09257703  
; Patent No. 6265538  
; GENERAL INFORMATION:  
; APPLICANT: Greene, Warner C.  
; APPLICANT: Lin, Xin  
; TITLE OF INVENTION: A NOVEL INHIBITOR OF THE INFLAMMATORY RESPONSE INDUCED  
; FILE REFERENCE: 30448.61USU1  
; CURRENT APPLICATION NUMBER: US/09/257,703  
; CURRENT FILING DATE: 1999-02-25  
; EARLIER APPLICATION NUMBER: 60/076,299  
; EARLIER FILING DATE: 1998-02-27  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 947  
; TYPE: PRT  
; ORGANISM: NF-KAPPA B INDUCING KINASE (NIK)  
US-09-257-703-1

Query Match 16.0%; Score 68; DB 4; Length 947;



QY 4 PFHCROIGSVYRLPL--RKGEVLP-----LPE-ANFPs----- 34  
125 PYHKRVESPV-LPVLVPRHNEFNQHSILVQFRNLSHNEPHMPONATFPDSFHQPNNNA 183  
QY 35 -PPL-PHKKHPLKPDNPFPQSVSESCPG 61  
184 PFPLSPNSPYPPSPASTYPPNSPASSGPG 212

## RESULT 8

US-08-840-767-48  
; Sequence 48, Application US/08840767B  
; Patent No. 6255464  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: Riggins, Gregory J.  
; APPLICANT: Thagalingam, Sam  
; TITLE OF INVENTION: MAD-Related Genes in the Human  
; FILE REFERENCE: 01107.05548  
; CURRENT APPLICATION NUMBER: US/08/840,767B  
; EARLIER FILING DATE: 1997-04-16  
; EARLIER APPLICATION NUMBER: 60/015,823  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 48  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-840-767-48

Query Match 15.3%; Score 65; DB 4; Length 465;  
Best Local Similarity 27.0%; Pred. No. 4, 6;  
Matches 24; Conservative 12; Mismatches 21; Indels 32; Gaps 6;

QY 4 PFHCROIGSVYRLPL--RKGEVLP-----LPE-ANFPs----- 34  
125 PYHKRVESPV-LPVLVPRHNEFNQHSILVQFRNLSHNEPHMPONATFPDSFHQPNNNA 183  
QY 35 -PPL-PHKKHPLKPDNPFPQSVSESCPG 61  
184 PFPLSPNSPYPPSPASTYPPNSPASSGPG 212

## RESULT 9

US-08-840-767-52  
; Sequence 52, Application US/08840767B  
; Patent No. 6255464  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: Riggins, Gregory J.  
; APPLICANT: Thagalingam, Sam  
; TITLE OF INVENTION: MAD-Related Genes in the Human  
; FILE REFERENCE: 01107.05548  
; CURRENT APPLICATION NUMBER: US/08/840,767B  
; EARLIER FILING DATE: 1997-04-16  
; EARLIER APPLICATION NUMBER: 60/015,823  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-840-767-52

Query Match 15.1%; Score 64; DB 4; Length 465;  
Best Local Similarity 27.0%; Pred. No. 6;

Matches 24; Conservative 12; Mismatches 21; Indels 32; Gaps 6;

QY 4 PFHCROIGSVYRLPL--RKGEVLP-----LPE-ANFPs----- 33  
125 PYHKRVESPV-LPVLVPRHNEFNQHSILVQFRNLSHNEPHMPONATFPDSFHQPNNNT 183  
QY 34 SEPL-PHKKHPLKPDNPFPQSVSESCPG 61  
184 PFPLSPNSPYPPSPASTYPPNSPASSGPG 212

## RESULT 10

US-09-096-776B-10  
; Sequence 10, Application US/09096776B  
; Patent No. 6270994  
; GENERAL INFORMATION:  
; APPLICANT: Miyazono, Kohel  
; APPLICANT: Kawabata, Masahiro  
; TITLE OF INVENTION: SMAD6 AND USES THEREOF  
; FILE REFERENCE: L0461/7038  
; CURRENT APPLICATION NUMBER: US/09/096,776B  
; CURRENT FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: US 60/049,990  
; PRIOR FILING DATE: 1997-06-13  
; PRIOR APPLICATION NUMBER: US 60/053,040  
; PRIOR FILING DATE: 1997-07-18  
; PRIOR APPLICATION NUMBER: US 60/066,173  
; PRIOR FILING DATE: 1997-11-18  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-096-776B-10

Query Match 15.1%; Score 64; DB 4; Length 465;  
Best Local Similarity 27.0%; Pred. No. 6;  
Matches 24; Conservative 12; Mismatches 21; Indels 32; Gaps 6;

QY 4 PFHCROIGSVYRLPL--RKGEVLP-----LPE-ANFPs----- 34  
125 PYHKRVESPV-LPVLVPRHNEFNQHSILVQFRNLSHNEPHMPONATFPDSFHQPNNNA 183  
QY 35 -PPL-PHKKHPLKPDNPFPQSVSESCPG 61  
184 PFPLSPNSPYPPSPASTYPPNSPASSGPG 212

## RESULT 11

US-08-999-774A-2  
; Sequence 2, Application US/08999774A  
; Patent No. 6274312  
; GENERAL INFORMATION:  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Seghezzi, Wolfgang  
; APPLICANT: Shanahan, Frances  
; APPLICANT: Lees, Emma M.  
; APPLICANT: McClanahan, Terrill K.  
; TITLE OF INVENTION: Intracellular Regulatory Molecules;  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/999,774A  
FILING DATE: 10-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/032,818  
FILING DATE: 11-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0646  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)852-9196  
TELEFAX: (650)496-1200  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1105 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-999-774A-2

Query Match 15.1%; Score 64; DB 4; Length 1105;  
Best Local Similarity 31.0%; Pred. No. 16;  
Matches 22; Conservative 6; Mismatches 35; Indels 8; Gaps 3;

QY 1 GPRPHCRQIGSYVRLP-LRKGEVLPLPEANFSPFLPHHK---HPLKPDNQPFQSVS 55  
Db 964 GNPQQAQHSGGRLAPLGAHGMHMQRPYRLMNMRRPRPGQIGPFSM 1023

QY 56 SESCGKFKSG 66  
Db 1024 --MFGHMPG 1031

RESULT 12  
US-08-301-162-16  
Sequence 16, Application US/08301162  
Patent No. 6022546  
GENERAL INFORMATION:  
APPLICANT: Knapp, Stefan  
APPLICANT: Ziegelmair, Robert  
APPLICANT: Kupper, Hans  
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The  
TITLE OF INVENTION: Preparation Thereof and the Use Thereof  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &  
Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/301,162  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,128  
FILING DATE:  
APPLICATION NUMBER: US 07/623,086  
FILING DATE: 06-DEC-1990  
APPLICATION NUMBER: DE P3940598.2  
FILING DATE: 08-DEC-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Flesher, Raz E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 02481.1005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4400  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-301-162-16

Query Match 15.0%; Score 63.5; DB 3; Length 267;  
Best Local Similarity 33.9%; Pred. No. 3.6;  
Matches 21; Conservative 4; Mismatches 30; Indels 7; Gaps 3;

QY 1 GPRPHCRQIGSYVRLP-LRKGEVLPLPEANFSPFLPHHKHPLKPDNQPFQSVSES 58  
Db 26 GPLSTHPSYSASYPNPSNPLHG--MPKPN--PVRPPGPHSVIAPNPPLGTPAS 80

QY 59 CP 60  
Db 81 MP 82

RESULT 13  
US-09-461-240-16  
Sequence 16, Application US/09461240  
Patent No. 6326008  
GENERAL INFORMATION:  
APPLICANT: Knapp, Stefan  
APPLICANT: Ziegelmair, Robert  
APPLICANT: Kupper, Hans  
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The  
TITLE OF INVENTION: Preparation Thereof and the Use Thereof  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &  
Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/461,240  
FILING DATE: 16-DEC-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/301,162  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/167,128  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/623,086  
FILING DATE: 06-DEC-1990  
APPLICATION NUMBER: DE P3940598.2  
FILING DATE: 08-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Flesher, Raz E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 02481.1005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4400  
TELEFAX: 202-408-4400

```

? INFORMATION FOR SEQ ID NO: 16:
?     SEQUENCE CHARACTERISTICS:
?         LENGTH: 267 amino acids
?         TYPE: amino acid
?         TOPOLOGY: linear
?     MOLECULE TYPE: peptide
?     SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-461-240-16

```

Query Match	15.0%;	Score 63.5;	DB 4;	Length 267;
Best Local Similarity	33.9%;	Pred. No. 3.6;		
Matches 21; Conservative	4;	Mismatches 30;	Indels 7;	Gaps 3;

QY	59 CP 60	81 MP 82
QY	1 GPRFHOQIGSVYLR--PLRKGVELRYREANFSPFLNNKRLRPDQRPPOSYS 58	26 GLSTYHSSSGASYRNPNSMLNG--MKREN--FVRRPRGGRSVIRNPRLDTPRAS 800
Db		

RESULT 14  
US-09-303-064-54  
; Sequence 54, Application US/09303064

```

GENERAL INFORMATION:
APPLICANT: MAINE, Gregory T.
APPLICANT: HUNT, Jeffery C.
APPLICANT: BROUENAC, Susan
APPLICANT: YIH-TSING, SHEU, Michael
APPLICANT: CHOYAN, Linda E.
APPLICANT: TYNER, Joan D.
APPLICANT: HOWARD, Lawrence V.
APPLICANT: PARBLEY, Stephen F.
APPLICANT: REMINGTON, Jack S.
APPLICANT: ARAUJO, Fausto
APPLICANT: SUZUKI, Yashuhiko
APPLICANT: LI, Shuli
TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
FILE REFERENCE: 6361.US.P1
CURRENT APPLICATION NUMBER: US/09/303,064
CURRENT FILING DATE: 1999-04-30
EARLIER APPLICATION NUMBER: 09/086,503
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 54
LENGTH: 398
TYPE: PRF
ORGANISM: Toxoplasma gondii
IS-09-303-064-54

```

Query Match	15.0%;	Score 63.5;	DB 4;	Length 398;
Best Local Similarity	33.9%;	Pred. No. 5.8;		
Matches 21; Conservative	4;	Mismatches 30;	Indels 7;	Gaps 3;

Oy	1	GRREHCISQISUYRL--RYKGEVLYLRANRFRFRLYNNKRYLRDQRPQSYSES	58
		:  :	
Db	174	GRLSYHPSYSGASTPNSNPLNG---MKREN--RVKRRFRGGFHSYTPNRPYLGTPAS	228
Oy	59	CP	60
Db	229	MP	230

RESULT 15  
US-09-086-503-54  
; Sequence 54, Application US/09086503A  
; Patent No. 6329157  
; GENERAL INFORMATION:

```

1  APPLICANT: MAINE, Gregory T.
2  APPLICANT: HUNT, Jeffrey C.
3  APPLICANT: BROJANAC, Susan
4  APPLICANT: JYH-TSING SHEU, Michael
5  APPLICANT: CHOYAN, Linda E.
6  APPLICANT: TYNER, Joan D.
7  APPLICANT: HOWARD, Lawrence V.
8  TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
9  FILE REFERENCE: 6361.US.-01
10 CURRENT APPLICATION NUMBER: US/09/086,503A
11 CURRENT FILING DATE: 1998-05-28
12 NUMBER OF SEQ ID NOS: 55
13 SOFTWARE: FastSeq for Windows Version 3.0
14 SEQ ID NO 54
15 LENGTH: 398
16 TYPE: PRF
17 ORGANISM: Toxoplasma gondii
18 US-09-086-503-54

```

Query Match	15.0%;	Score 63.5;	DB 4;	Length 398;
Best Local Similarity	33.9%;	Pred. No. 5.8;		
Matches	21;	Conservative	4;	Mismatches 30;
				Indels 7;
				Gaps 3;

```

0y 1 GPRFHCROIGSVRLP--PLRKGVELDLEPANEPSFLDHHKHPKPDNOPEQSVSES 58
    || : | | | | | | | | | | | | | | | | | | | | | | | | | |
174 GPLSYHPSSGASYPNPSPNLHG--MKPEN--PVRPPPGGFHPSVIANPPYPLGTPPAS 228
Db

```

QY	59	CP	60
Db	229	MP	230

Search completed: May 24, 2002, 16:48:58  
Job time: 283 sec

RESULT 15  
US-09-086-503-54  
; Sequence 54, Application US/09086503A  
; Patent No. 6329157  
; GENERAL INFORMATION:  
3

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

Run on: May 24, 2002, 16:44:00 ; Search time 136.41 Seconds  
(without alignments)  
61.070 Million cell updates/sec

```
Title: US-09-730-379E-1
Perfect score: 424
Sequence: 1 GRRPFHCRQIGSVYRLPLR.....SESCPGKRSQFPQVSMFTT 75
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database : A\_Geneseq\_032802:\*

1:	A_Geneseq.022802.*
2:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1980.DAT *
3:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1981.DAT *
4:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1982.DAT *
5:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1983.DAT *
6:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1984.DAT *
7:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1985.DAT *
8:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1986.DAT *
9:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1987.DAT *
10:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1988.DAT *
11:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1990.DAT *
12:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1991.DAT *
13:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1992.DAT *
14:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1993.DAT *
15:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1995.DAT *
16:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1995.DAT *
17:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1996.DAT *
18:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1997.DAT *
19:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1998.DAT *
20:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA1999.DAT *
21:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA2000.DAT *
22:	/SDSI/gcgdata/hold-genseq/geneseqp-emb1/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	77.5	18.3	158	21	AAV54568	A synthetically tr
2	77.5	18.3	745	22	AAU25461	Human mdt protein
3	75	17.7	130	22	AAO10033	Human polypeptide
4	74	17.5	446	22	ABB70063	Drosophila melanog
5	74	17.5	530	22	ABB65378	Drosophila melanog
6	73	17.2	607	22	AAU32109	Novel human secret
7	72.5	17.1	584	20	AAV28992	Mouse Tcf-3 (mtcf
8	72	17.0	554	22	ABB71099	Drosophila melanog
9	71.5	16.9	67	22	ABB57756	Drosophila melanog
10	71.5	16.9	179	22	ABG07476	Novel human diagn
11	71.5	16.9	485	19	AAW45913	Mouse Ckr serine/v

12	71	16.7	204	21	AAB41770	Human OREF ORF1534
13	70.5	16.6	136	22	AAO01200	Human polypeptide
14	70.5	16.6	1729	22	ABB82538	Drosophila melanog
15	70	16.5	445	22	AA411008	Human polypeptide
16	69.5	16.4	1232	22	AAU32866	Human human secret
17	69	16.3	174	21	AAB51684	Human OREF ORF1448
18	69	16.3	373	22	AA895352	Human protein sequ
19	68	16.0	644	22	ABG18134	Novel human diagn
20	68	16.0	947	18	AA412402	Novel human diagn
21	68	16.0	947	20	AAV31665	Amilo acid sequenc
22	68	16.0	947	20	AAV31667	NF-kB inducing kin
23	68	16.0	947	20	AAV31668	NIK kinase-deficie
24	68	16.0	947	20	AAAM82497	Human NIK protein
25	68	16.0	947	20	AAAM81564	Human NF-kB-induc
26	68	16.0	947	20	AAAM81561	Human NF-kB-induc
27	68	16.0	1373	22	ABBG6368	Drosophila melanog
28	67.5	15.9	511	19	AAW72141	HSV-2 strain SB5 C
29	67.5	15.9	511	19	AAW72073	HSV-2 strain SB5 C
30	66.5	15.7	73	20	AAW74149	Human prostate tun
31	66.5	15.7	179	22	AAAB36630	Human FLEXHF-52 p
32	66.5	15.7	351	22	AAAB52310	Chicken zyxine fra
33	66.5	15.7	481	19	AAAW2161	HSV-2 strain SB5 C
34	66.5	15.7	542	22	AAAM5304	Novel human diagn
35	66	15.6	172	22	ABG23495	Chicken zyxine.. G
36	66	15.6	526	21	AAAB43199	Novel human diagn
37	66	15.6	729	17	AAAB89275	Human OREF ORF29635
38	65.5	15.4	528	22	ABG14691	yeast coagulation
39	65	15.3	330	22	ABBB30894	Novel human diagn
40	65	15.3	330	22	ABBB36076	Peptide #3545 encc
41	65	15.3	330	22	AAAM68858	Peptide #3582 encc
42	65	15.3	330	22	AAAM69245	Human brain expres
43	65	15.3	330	22	AAAM17074	Human bone marrow
44	65	15.3	330	22	AAAM29568	Peptide #3508 encc
45	65	15.3	330	22	AAAM29568	Peptide #3605 encc
	65	15.3	330	22	AAAM04775	Peptide #3457 encc

## ALIGNMENTS

RESULT	1
AA1545568	
ID	AA1545468 standard; Protein; 158 AA.
XX	
XX	AA1545568;
XX	
DT	25-APR-2000 (first entry)
XX	
DE	A synthetically truncated C hordein prot
XX	
KW	C hordein gene; glutenin; seed storage
KW	noodle; breakfast cereal; snack food; c
KW	film; coating; adhesive; building materi
XX	
XX	
OS	Hordeum vulgare.
OS	Synthetic.
XX	
XX	
FT	Key
FT	Misc-difference 20
FT	Location/Qualifiers
FT	/note= "encoded by CCG"
FT	Misc-difference 105
FT	/note= "encoded by TTC"
FT	Misc-difference 106
FT	/note= "encoded by CCC"
FT	Misc-difference 107
FT	/note= "encoded by TGG"
FT	Misc-difference 108
FT	/note= "encoded by CAA"
FT	Misc-difference 113
FT	/note= "encoded by TTT"
FT	Misc-difference 114
FT	/note= "encoded by CCC"
FT	Misc-difference 115
FT	/note= "encoded by CAG"

FT	Misc-difference	116	/note= "encoded by CCC"
FT	Misc-difference	117	/note= "encoded by CAA"
FT	Misc-difference	118	/note= "encoded by GAA"
FT	Misc-difference	119	/note= "encoded by CCA"
FT	Misc-difference	120	/note= "encoded by ATT"
PV	Wo200002914-AL.		
PD	20-JAN-2000.		
XX			
PF	12-JUL-1999;	99WO-AU00563.	
XX			
PR	10-JUL-1998;	98AU-0004604.	
PA	(CSTR ) COMMONWEALTH SCI & IND RES ORG.		
PA	(GOOD-) GOODMAN FIELDER LTD.		
XX	(LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.		
PI	Appels R, Morell M, Bekes F, Tamas L;		
DR	WPI: 2000-147597/13.		
N-PSDB:	AAZ45987.		
PT	Modifying glutenin or seed-storage protein for preparing foodstuffs,		
PT	films, coatings, packing materials, adhesives and building materials -		
XX			
PS	Disclosure: Flg 3; 76pp; English.		
XX			
CC	The present sequence represents a synthetically truncated barley C		
CC	hordein protein, which is a glutelin. The protein is designated		
CC	ANG-deltaacy7S7ys236, and is modified, according to the method of		
CC	the invention. The specification describes a method for producing a		
CC	modified glutenin or seed storage protein, by adding to the protein a		
CC	domain that confers the ability to incorporate into gluten, or to bind		
CC	a ligand or other macromolecule. The domain can be any domain that will		
CC	bind ligands that may be useful in food preparation or in food		
CC	composition, e.g. a domain that binds lipids or starch. The method is		
CC	used for producing modified glutenins. Glutenins and seed-storage protein		
CC	are useful for preparing food products such as leavened or unleavened		
CC	breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries		
CC	and foods containing flour based sauces. Glutenins and seed-storage		
CC	protein are also useful for preparing nonfood products such as films,		
CC	coatings, adhesives, building materials and packaging materials. Grain		
CC	or parts of grain containing the modified glutenin and seed-storage		
CC	protein is useful for preparing food products. The modified glutenins and		
CC	seed-storage proteins are also useful as modifiers of food products in		
CC	food industry.		
SSO	Sequence    158 AA;		
Query Match	18.3%; Score 77.5; DB 21; Length 158;		
Best Local Similarity	27.7%; Pred. No. 0.53;		
Matches	26; Conservative 11; Mismatches 28; Indels 29; Gaps 5		
OY	3 RFEGCRQ-----ICSVRLP-----PLKRGVLDPREANFP-----SPFLRN 39		
Dd	:      :       :       :       :       :       :       :		
	qRfHeRqyfyRpLeelfpqyflPrPlrpqpqrPqlrPpqqrPfqPpqPfpPpqPq 97		
OY	40 HKNRPLDQRFRVSSESCRGKFKSGRFQVMNF 73		
Dd	:        :   :   :   :   :   :   :   :   :   :   :   :		
	eRpRpqr qqrPpqPpqPfqPpq----pqqlf 125		
RESULT	2		
ID	AAU25461 standard; Protein; 745 AA.		
XX			

AC	AAU25461;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human mdlt protein from clone LI:243660.4:2000MAY01.
XX	
KW	Human; molecules for disease detection and treatment; mdlt;
KW	Antiartherosclerotic; hepatotropic; antipsoriatic; cytostatic;
KW	immunosuppressive; antidiabetic; antiasthmatic; neuroprotective;
KW	osteoprotic; antiarthritic; cell proliferative disorder;
KW	arteriosclerosis; cirrhosis; psoriasis; cancer; adenocarcinoma;
KW	leukaemia; breast cancer; autoimmune disorder; AIDS;
KW	acquired immunodeficiency syndrome; Addison's disease;
KW	diabetes mellitus; asthma; multiple sclerosis; osteoarthritis.
KX	
OS	Homo sapiens.
PN	WO200162922-A2.
XX	
PD	30-AUG-2001.
XX	
PF	21-FEB-2001; 2001MO-USO5896.
XX	
PR	24-FEB-2000; 2000US-0185213.
PR	16-MAY-2000; 2000US-0205232.
PR	17-MAY-2000; 2000US-0205285.
PR	17-MAY-2000; 2000US-0205286.
PR	17-MAY-2000; 2000US-0205287.
PR	17-MAY-2000; 2000US-0205323.
PR	17-MAY-2000; 2000US-0205324.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Panzer SR, Spiro PA, Banyille SC, Shah P, Chalup MS, Chang SC;
PI	Chen A, D'Sa SA, Amshley S, Dahl CR, Dam TC, Daniels SE;
PI	Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI	Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Dafo A;
PI	Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI	Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
DR	WPI: 2001-570631/64.
N-PSDB:	AAS42513.
XX	
PT	New disease detection and treatment molecule polynucleotides and
PT	polypeptides, useful for diagnosis and treatment of arteriosclerosis,
PT	cirrhosis, psoriasis, cancer, autoimmune disorders, diabetes mellitus,
PT	asthma and multiple sclerosis -
XX	
PS	Claim 27; Page 172-173; 183pp; English.
XX	
CC	The invention relates to novel human molecules for disease
CC	detection and treatment (mdlt proteins) and the polynucleotides encoding
CC	them. The MDLT polynucleotides and polypeptides are useful for diagnostic
CC	and therapeutic purposes e.g. to diagnose and treat cell proliferative
CC	disorders (e.g. arteriosclerosis, cirrhosis and psoriasis) cancers (e.g.
CC	adenocarcinoma, leukaemia and breast cancer) autoimmune disorders
CC	(e.g. acquired immunodeficiency syndrome (AIDS) and Addison's disease)
CC	diabetes mellitus, asthma, multiple sclerosis, osteoarthritis, and many
CC	more diseases given in the specification. The present sequence
CC	represents an mdlt protein of the invention.
XX	
SO	Sequence 745 AA;
Query Match	18.3%; Score 77.5; DB 22; Length 745;
Best Local Similarity	31.6%; Pred. No. 3.2;
Matches 24; Conservative	7; Mismatches 28; Indels 17; Gaps 3
OY	1 GPRPHCRIGV-----YRLRPL-RKGEVLRLP-----EANPPSPDLPHKKP 43
db	620 grrlglilgppaalgkagaitqvyllprlpqqlqvaparpargytkaaaamrptchhhp 679
OY	44 LKPDNOPEPOSVSESC 59



Db 680 fhppcghfhgrspgpc 695

## RESULT 3

AA010033  
ID AA010033 standard; Protein: 130 AA.  
XX  
AC AA010033;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 23925.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PE 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSEQ) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB: AAI89964.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 23925; 13999P + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 130 AA:

Query Match 17.7%; Score 75; DB 22; Length 130;

Best Local Similarity 39.1%; Pred. No. 0.8; Mismatches 25; Conservative 5; Indels 14; Gaps 20; Gaps 5;

QY 17 PLRKGEVLPRLPEANFSPF-----PLP-----HHKHPD---KPDNPFPOSVSESCPGK-F 63  
DB 9 pprirgsktpppprkilfprrfkpprkpkcxknkoplqgpkkkknkrrp-----pgrtf 61  
QY 64 KSGF 67  
DB 62 kgkf 65

## RESULT 4

ABB70063  
ID ABB70063 standard; Protein: 446 AA.  
XX  
AC ABB70063;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 36981.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE) PE CORP NY;  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB: ABL14166.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure: SEQ ID NO 36981; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (AB101840-ABL16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 446 AA:

Query Match 17.5%; Score 74; DB 22; Length 446;

Best Local Similarity 38.1%; Pred. No. 4.3; Mismatches 24; Conservative 3; Indels 28; Gaps 8; Gaps 3;

QY 2 PRPFHCROIGSVYRLPLRKGEVLP---PEANFSPRLPHNHKHPKPDNPFPOSVSRS 58  
DB 294 prp---qpsey-lpppenevtrpqrptarvpeypppppprpprtyqprppppr 348  
QY 59 CPG 61  
DB 349 apq 351

## RESULT 5

ABB66378  
ID ABB66378 standard; Protein: 530 AA.  
XX  
AC ABB66378;  
XX  
DT 26-MAR-2002 (first entry)

```

XX DE Drosophila melanogaster polypeptide SEQ ID NO 25926.
XX PF
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PA WO200171042-A2.
XX PN
XX PD 27-SEP-2001.
XX PF 23-MAR-2001: 2001WO-US09231.
XX PR 23-MAR-2000: 2000US-191637P.
XX PR 11-JUL-2000: 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR MPI: 2001-656860/75.
XX DR N-PSDB; ABL10481.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 25926; 21bp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX CC sequences (AB11840-AB116175) and the encoded proteins
XX CC (ABBS7737-ABBS7072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 530 AA:

Query Match 17.5%; Score 74; DB 22; Length 530;
Best Local Similarity 43.2%; Pred. No. 5.3;
Matches 16; Conservative 3; Mismatches 14; Indels 4; Gaps 2;

QY 21 KGEVLPPEANFPSPFLPHHKHLKLPDNOFPQSVSE 57
   : : : : : : : : : : : : : : : : : : : :
Db 129 egaslpbpe--pkhpmlekhkplpsltlpi-asvke 161

RESULT 6
AAU32109
ID AAU32109 standard; Protein; 607 AA.
XX AC AAU32109;
XX DE 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #2600.
XX KW Human; vaccination; gene therapy; nutritional supplement;
XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.

```

```

XX XX 16-APR-2001: 2001WO-US08656.
XX XX 18-APR-2000: 2000US-0552929.
XX PR 26-JAN-2001: 2001US-0770160.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR MPI: 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20: Page 556; 765bp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising
XX CC the nucleic acids encoding the polypeptides and cells genetically
XX CC engineered to express them are also useful for producing the proteins.
XX CC The proteins are useful in genetic vaccination, testing and
XX CC therapy, and can be used as nutritional supplements. They may be used to
XX CC increase stem cell proliferation; to regulate haematopoiesis; and in
XX CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX CC immune suppression and/or stimulation; as anti-inflammatory agents; and
XX CC in treatment of leukaemias. AA029510-AA03304 represent the amino acid
XX CC sequences of novel human secreted proteins of the invention.
XX SQ Sequence 607 AA:

Query Match 17.2%; Score 73; DB 22; Length 607;
Best Local Similarity 32.3%; Pred. No. 8;
Matches 30; Conservative 6; Mismatches 33; Indels 24; Gaps 6;

QY 2 PREPFCNQISGYRLRP-LRKGEVLP-----LPENFPSPFLPHHK 41
   : : : : : : : : : : : : : : : : : : : :
Db 461 pmlhprfpfs-srlp91lggeydpkntlfmlgdpisslpwvlgefpqspfp-plet 518

QY 42 HPLKPDNQPPQSVS-ESCPGKFKSGFPQVSMF 73
   : : : : : : : : : : : : : : : : : : : :
Db 519 halnpswpplsrldnpqscpqa-egppprtrtf 550

RESULT 7
AAV28992
ID AAV28992 standard; Protein; 584 AA.
XX AC AAV28992;
XX DE 05-OCT-1999 (first entry)
XX DE Mouse Tcf-3 (mtcf-3) protein.
XX KW Mammalian; transcription factor; Tcf-3; Tcf-4; beta-catenin; mouse;
XX KW binding domain; Wnt signalling; Wnt/Wingless cascade; anticancer drug;
XX KW transactivation; fetal cancer; testis cancer; ovary teratoma; mammary;
XX KW myeloblastoma; medulloblastoma; intestinal cancer; response element.
XX OS Mus sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 5
XX FT /note= "encoded by GGT"
XX FT /note= "encoded by TCT"
XX FT Misc-difference 225
XX FT /note= "encoded by TCT"

```

XX EP939122-A1.  
 PN  
 XX  
 PD 01-SEP-1999.  
 XX  
 XX 27-FEB-1998; 98EP-0200621.  
 PF  
 XX 27-FEB-1998; 98EP-0200621.  
 PR  
 XX 27-FEB-1998; 98EP-0200621.  
 XX  
 PA (UUT-) RIKSUNIV UTRECHT.  
 DR  
 XX WPI: 1999-471050/40.  
 DR N-PSDB: AAX99684.  
 XX  
 PT New mouse transcription factors, useful for diagnosing and treating  
 PT cancers and screening for anticancer drugs  
 PS  
 XX Disclosure; Fig 1; 56pp; English.  
 CC The invention relates to mammalian transcription factors Tcf-3 or Tcf-4  
 CC protein or their derivatives. The tcf proteins of the invention comprise  
 CC a beta-catenin binding domain and are capable of complexing with beta-  
 CC catenin in response to signaling through the Wnt/Wingless cascade. The  
 CC Tcf-3 and Tcf-4 genes, proteins and antibodies against the proteins may  
 CC be used to screen for candidate anticancer drugs by identifying compounds  
 CC which interact with beta-catenin/Tcf-3 or beta-catenin/Tcf-4 complexes  
 CC and alter their transactivation activity. Compounds which interact with  
 CC the Tcf-3 and Tcf-4 proteins and interfere with their transcriptional  
 CC activation, identified using these screens, may be used to treat fetal  
 CC cancer, cancers with stem cell phenotypes, testis or ovary teratomas,  
 CC myeloblastomas, medulloblastomas, mammary or intestinal cancers. The  
 CC present sequence represents a mouse Tcf-3 (mTcf-3) protein sequence.  
 XX  
 SQ Sequence 584 AA;  
 Query Match 17.1%; Score 72.5; DB 20; Length 584;  
 Best Local Similarity 29.8%; Pred. No. 8.7;  
 Matches 25; Conservative 9; Mismatches 33; Indels 17; Gaps 4;  
 QY 2 PRPHCRIGSVYRLPKGEV-LPL---PEANFSPFLP---HHKHPKLPDN--- 48  
 Db 219 prphcpelispypispaavqgipblwlvpgqgpmyslppgqftrhpalamaas 278  
 QY 49 ----QPFQSVSESCPGKFKSGFP 68  
 Db 279 slvsstrfpmvapanpglptsgip 302  
 RESULT 8  
 ABB71099  
 ID ABB71099 standard; Protein: 554 AA.  
 AC ABB71099;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 40089.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 EN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 PD  
 PF 23-MAR-2001; 2001WO-US09231.  
 PF  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX

PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB: ABL15202.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS  
 XX Disclosure; SEQ ID NO 40089; 21pp + Sequence listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 554 AA;  
 Query Match 17.0%; Score 72; DB 22; Length 554;  
 Best Local Similarity 30.5%; Pred. No. 9.3;  
 Matches 25; Conservative 9; Mismatches 32; Indels 16; Gaps 5;  
 QY 1 GPR--PF---HCROIGSVYRLPKGEV-LPLPEANFSPFLPHKHPKLPDNQFP-- 52  
 Db 441 gphngprfvehgngpqlfgrfdvqrgsegegrfvehqprfdrhian---rppaeifrp 496  
 QY 53 ---QSVSES-CPGKFKSGFPQV 70  
 Db 497 rrsrlsrshckekgesfpkv 518  
 RESULT 9  
 ABB57756  
 ID ABB57756 standard; Protein: 67 AA.  
 AC ABB57756;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 60.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 EN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 PD  
 PF 23-MAR-2001; 2001WO-US09231.  
 PF  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB: ABL01859.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT	genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PS	Disclosure; SEQ ID NO 60; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU101840-ABL16175) and the encoded proteins (ABU57737-ABR72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
SC	Sequence 67 AA:
OY	Query Match 16.9%; Score 71.5; DB 22; Length 67; Best Local Similarity 35.0%; Pred. No. 0.91; Mismatches 19; Indels 15; Gaps 4 Matches 21; Conservative 5;
DB	26 PLEPANEFSPFLPH-----HKHPLKPDNQPP--QSVSESCPGRKS----GEPQVS 71   :   :                                     4 pppdqgfiyhphlpthttbpshhp-hpshpbphqnpqlqppgfmpfdmsfgrs 62
RESULT 10	
ID	ABG07476 standard; Protein: 179 AA.
XX	
AC	ABG07476;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #7467.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
RW	
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
PC	
XX	11-OCT-2001.
PD	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73. N-PSTDB; AAS71663.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PT	
PS	Claim 20; SEQ ID NO 37835; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques
CC	

CC	to restore normal activity of (II) or to treat disease states involving	CC
CC	(II). (II) is useful for generating antibodies against it, detecting or	CC
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	CC
CC	a food supplement. (II) and its binding partners are useful in medical	CC
CC	imaging of sites expressing (II). (I) and (II) are useful for treating	CC
CC	disorders involving aberrant protein expression or biological activity.	CC
CC	The polypeptide and polynucleotide sequences have applications in	CC
CC	diagnostics, forensics, gene mapping, identification of mutations	CC
CC	responsible for genetic disorders or other traits to assess biodiversity	CC
CC	and to produce other types of data and products dependent on DNA and	CC
CC	amino acid sequences. ABG00010-ABG30377 represent novel human	CC
CC	diagnostic amino acid sequences of the invention.	CC
CC	Note: The sequence data for this patent did not appear in the printed	CC
CC	specification, but was obtained in electronic format directly from WIPO	CC
CC	at ftp.wipo.int/pub/published_pct_sequences.	CC
xx		xx
SO	Sequence 179 AA:	SO
Query Match	16.9%; Score 71.5; DB 22; Length 179;	
Best Local Similarity	27.4%; Pred. No. 2.8;	
Matches 26; Conservative 11; Mismatches 33; Indels 25; Gaps 5		
Qy	3 REFRCRQIGSVYRLP-PLRKG--EVLPL-----PEANPPSPFLPHHKPLKRDNQ- 49	
Db	53 Ipcrappvdtlyrllspilrqlaghtlplrlpsftrspillyps-pprhahylpsps 111	
Qy	50 -----PFQPSVSESCPGKFKSGPGQYSMF 73	
Db	112 mrrymaltavpyrlytsplqlafpfhlsrtpqlqlf 146	
RESULT 11		
AA049913		
ID	AA049913 standard; Protein; 485 AA.	
XX	AA049913;	
XX	20-JUL-1998 (first entry)	
DT		
XX	Mouse CLK serine/threonine kinase mCLK3.	
DE		
XX	mCLK3: CLK: serine/threonine kinase; protein kinase; LAMMER kinase;	
KW	signal transduction; cancer; contraceptive; mouse; therapy;	
KW	diagnosis.	
XX	Mus musculus.	
OS		
XX		
XX	Key Location/Qualifiers	
FH	23..41	
FT	/note="nuclear localisation domain"	
FT	Domain 151..469	
FT	/note="catalytic domain"	
FT	Peptide 377..382	
FT	/note="LAMMER motif"	
XX		
PN	W09748723-A2.	
XX		
XX	24-DEC-1997.	
PD		
XX	17-JUN-1997;	
PF	97WO-1B00946.	
XX		
PR	19-DEC-1996; 96US-0034286.	
PR	17-JUN-1996; 96US-0019629.	
PR	09-AUG-1996; 96US-0023485.	
PR	13-NOV-1996; 96US-0030860.	
PR	15-NOV-1996; 96US-0030964.	
XX		
PA	(PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.	
XX		
PI	Aoki N, Chen Z, Kharitonkov AI, Kim YW, Nayler O;	
PI	Ullrich A, Wang HY;	





XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB: AAI60164.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -

PS Example 2: SEQ ID NO 5939; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 445 AA;

Query Match 16.5%; Score 70; DB 22; Length 445;

Best Local Similarity 29.4%; Pred. NO. 12;  
 Matches 23; Conservative 5; Mismatches 27; Indels 28; Gaps 3;

OY 2 PRPFHCRQIGSYR-----LPPLRKG-----EVLPLPEANPPSPPLPHHKKHPLKP 46  
 DB 313 prpctshwpgvgyqfvtvpgvypipplissmprnlpplpplpseftlpsfpl----- 363  
 OY 47 DNQPPQSVSESCPGKFKSGFPQVS 71  
 DB 364 ---vpesssaassgellisslpts 384

Search completed: May 24, 2002, 16:48:02  
 Job time: 242 sec

**THIS PAGE BLANK (USPTO)**



RESULT	2	
Q99PS7		
ID	Q99PS7	PRELIMINARY; PRT; 515 AA

AC Q99PS7;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN 2.  
GN RNHRG2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB055896; BAB33093.1; .  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; cystatin; 1.  
DR SMART: SM00043; CY; 2.  
SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 67.1%; Score 213.5; DB 11; Length 515;  
Best Local Similarity 71.2%; Pred. No. 1.6e-20;  
Matches 42; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

OY 1 ASFRVDRIEVARVARGEGT-YFVDFSVRNCPRHHPRHNVGFCRADLFYDVEALDL 58  
DB 171 ASFRVERAEKRVIRGGERSTYFIERSVRCSTQHFPRHPVFGFCRAVLSTIEASDL 229

RESULT 3  
ID Q9ESB2 PRELIMINARY; PRT; 510 AA.  
AC Q9ESB2;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN.  
GN HRC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LEWIS;  
RA MEDLINE=20307726; PubMed=10849117;  
RA Hulet M.D., Parish C.R.;  
RT "Murine histidine-rich glycoprotein: cloning, characterization and  
RT cellular origin";  
RL Immunol. Cell Biol. 78:280-287(2000).  
DR EMBL: AF194029; AAG28417.1; .  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; cystatin; 1.  
DR SMART: SM00043; CY; 2.  
SQ SEQUENCE 510 AA; 57581 MW; 508B6E06AA2ED58E CRC64;

Query Match 64.6%; Score 205.5; DB 11; Length 510;  
Best Local Similarity 69.5%; Pred. No. 1.9e-19;  
Matches 41; Conservative 4; Mismatches 13; Indels 1; Gaps 1;

OY 1 ASFRVDRIEVARVARGEGT-YFVDFSVRNCPRHHPRHNVGFCRADLFYDVEALDL 58  
DB 171 ALFRVERAEKRVIRGGERSTYFIERSVRCSTQHFPRHPLVFGFCRAVLSTIEASDL 229

RESULT 4  
O99PS8 PRELIMINARY; PRT; 525 AA.  
ID Q99PS8

AC Q99PS8;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN 1.  
GN RNHRG1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB055895; BAB33092.1; .  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; cystatin; 1.  
DR SMART: SM00043; CY; 2.  
SQ SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match 64.3%; Score 204.5; DB 11; Length 525;  
Best Local Similarity 67.8%; Pred. No. 2.7e-19;  
Matches 40; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

OY 1 ASFRVDRIEVARVARGEGT-YFVDFSVRNCPRHHPRHNVGFCRADLFYDVEALDL 58  
DB 171 ASFRVERAEKRVIRGGERSTYFIERSVRCSTQHFPRHPVFGFCRAVLSTIEASDL 229

RESULT 5  
ID Q9TS85 PRELIMINARY; PRT; 214 AA.  
AC Q9TS85;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN-FACTOR XIIIA SUBSTRATE (FRAGMENTS).  
GN Bos taurus (Bovine).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94220160; PubMed=7909439;  
RA Halkier T., Andersen H., Vestergaard A., Magnusson S.;  
RT "Bovine histidine-rich glycoprotein is a substrate for bovine plasma  
RT factor XIIa";  
RL Biochem. Biophys. Res. Commun. 200:78-82(1994).  
DR InterPro: IPR000010; Cystatin.  
DR Pfam: PF00031; cystatin; 1.  
FT NON\_TER 1 1  
FT NON\_CONS 23 24  
FT NON\_CONS 91 92  
FT NON\_CONS 120 121  
FT NON\_CONS 180 181  
FT NON\_CONS 198 199  
FT NON\_TER 214 214  
SQ SEQUENCE 214 AA; 23982 MW; B8989492D6097A35 CRC64;

Query Match 63.1%; Score 200.5; DB 6; Length 214;  
Best Local Similarity 81.8%; Pred. No. 3.3e-19;  
Matches 36; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

OY 16 GGECT-YFVDFSVRNCPRHHPRHNVGFCRADLFYDVEALDL 58  
DB 121 GGECTSYFDFSVRNCSSHHFPRHSHIFGFCRADLFYDVEASDL 164

RESULT	6		
099PS6			
ID	099PS6	PRELIMINARY;	PRT; 525 AA.
AC	099PS6;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)		
DE	HISTIDINE-RICH GLYCOPROTEIN.		
GN	MMRG.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C; TISSUE=LIVER;		
RA	Matabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,		
RA	Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;		
RT	"Molecular diversity of mammalian histidine-rich glycoprotein.";		
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AB055897; BAB33094.1; -		
DR	InterPro; IPR00010; Cystatin.		
DR	Pfam; PF00031; cystatin; 1.		
DR	SMART; SM00043; CY; 2.		
SQ	SEQUENCE 525 AA; 59090 MW; AB3E93A439CFB3AC CRC64;		

	Query Match	61.8%	Score 196.5;	DB 11;	Length 525;
	Best Local Similarity	66.1%;	Pred. No. 3.3e-18;		
Matches	Conservative	39;	Mismatches	14;	Indels 1; Gaps 1
OY	1 ASFRDRLERPARVGGEST-YVDPSVRNCRPHRPPNPVFGRDALFDVEALDL	58			
DG	171 ASFRERERVIRAGGETNTIYFSMNCSTQHPRPSPDYFGRRALLASTVISFLSD	229			

RESULT	7		
Q9ESB3			
ID	Q9ESB3	PRELIMINARY;	PRT; 525 AA.
AC	Q9ESB3;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HISTIDINE-RICH GLYCOPROTEIN.		
GN	HRG.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid:10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN:129;		
RX	MEDLINE:20307726; PubMed:10849117;		
RA	Hulet M.D., Parish C.R.;		
RT	"Murine histidine-rich glycoprotein: cloning, characterization and		
RL	cellular origin.";		
RL	Immunol. Cell Biol. 78:280-287(2000).		
DR	EMBL; AF194028; AAC28416.1; -		
DR	InterPro; IPR000010; Cystatin.		
DR	Pfam; PF000031; Cystatin; 1.		
DR	SMART; SM00043; CY; 2.		
QO	SEQUENCE 525 AA; 39132 MW; 6E55F2A439CFB123 CRC64;		

	Query Match	Score	DB	length
Best Local Similarity	66.1%	Pred. No. 3	3e-18	
Matches	39	Conservative	5	Mismatches 14; Indels 1; Gaps 1
Qy	1	ASFRDRIERVARVGGEGT	-YFVDSVANCRRHHFPRHPNVFGCRADLFDEVALDL	58
Db	171	ASFPERERERTVRAAGGETNTNYVFESMNCSTQHPRSPPLVFGCRALLSTSIETSDL		229

RESULT	8			
ID	099PS5	PRELIMINARY;	PRT;	525 AA.
AC	099PS5;			
DT	01-JUN-2001 (TREMBLrel, 17, Created)			
DT	01-JUN-2001 (TREMBLrel, 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, Last annotation update)			
DE	HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).			
GN	MMRG.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
	[1]			
RP	SEQUENCE FROM N.A.			
RA	Tsuchida N., Wakabayashi S., Jahnhen-Dechent W., Koide T.;			
RT	"Structure of mouse histidine-rich glycoprotein gene.";			
RL	submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.			
	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER, NORMAL, 5 MONTH OLD MALE MOUSE.;			
RA	Strausberg R.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB055898; BAB33095.1; -			
DR	EMBL; BC011168; AAH11168.1; -			
DR	InterPro; IPR000010; Cystatin.			
DR	Pfam; PF00031; cystatin.1.			
DR	SMART; SM00043; CY; 2.			
SO	SEQUENCE 525 AA; 59162 MW; AB3E93AA39CFB126 CRC64;			

Query Match	61.8%	Score	196.5	DB	11	Length	525
Best Local Similarity	66.1%	Pred. No.	3	3e-18			
Matches	39	Conservative	5	Mismatches	14	Indels	1
				Gaps			1
QY	1	ASFRDRIERARVVGEGST	-YFVDESANVCNCRHHPPRRHNNVFGCRADLFYDVEALD	58			
DB	171	ASFRERERARVRRGEGTNNYVEESMNCSTQHPRRSLPVGFCRALLSTISFLSD	229				

RESULT	9		
09U5C0		PRELIMINARY:	PRT: 339 AA.
AC	09U5C0;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL 38.9 KDA PROTEIN.		
GN	F43C11.9.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	None;		
RT	"Genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium."		
RL	Science 282:2012-2018(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Gattung S., Du H., Scheel P., Hawrysko C.;		
RT	"The sequence of C. elegans cosmid F43C11."		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Waterson R.;		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.		

## RESULT 13

Q39046  
ID 039046 PRELIMINARY: PRT; 622 AA.  
AC 039046:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CER1-LIKE PROTEIN.  
GN CER1-LIKE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. MASSILEVSKAYA;  
RA Deng M.D., Peng S., Lemieux B.;  
RT "Genomic (Accession No. X95964) and cDNA (Accession No. X95965)  
sequences of the CER1-like gene of Arabidopsis thaliana derived from a  
RT plant DNA/T-DNA insertion junction. (PGR96-019).";  
RL Plant Physiol. 110:1436-1436(1996).  
DR EMBL: X95964; CAA65199.1;  
DR InterPro: IPR001541; Sterol\_desat.  
DR Pfam: PF01598; Sterol\_desat; 1.  
SQ SEQUENCE 622 AA; 71457 MW; DDEF0FALC6967493 CRC64;

Query Match 19.5%; Score 62; DB 10; Length 622;  
Best Local Similarity 41.3%; Pred. No. 6.8;  
Matches 19; Conservative 3; Mismatches 12; Indels 12; Gaps 3;

OY 6 DRIERAVRGGEGTYFVDFSVNCRPHRPNVFGCRADLFY 51  
Db 513 DGIENEDQMKPKETLFPSP-----HP--PNKL---RKOCFY 546

RESULT 14  
09L173 PRELIMINARY: PRT; 452 AA.  
AC 09L173:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CHLOROPLAST NUCLEOID DNA BINDING PROTEIN-LIKE, NUCCELLIN-LIKE  
DE PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE-20363099; PubMed-10907853;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
RT TAC and BAC clones."  
RL DNA Res. 7:217-221(2000).  
DR EMBL: AP001313; BAB03090.1; -  
DR HSSP: P00797; ZREN.  
DR InterPro: IPR001969; Asp\_protease.  
DR InterPro: IPR001461; Pepsin.  
DR InterPro: IPR001230; Prenyltn.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
DR PROSITE: PS00294; PRENYLATION; UNKNOWN\_1.  
SQ SEQUENCE 452 AA; 49374 MW; DDCFO9D45B566410 CRC64;

Query Match 19.2%; Score 61; DB 10; Length 452;  
Best Local Similarity 25.8%; Pred. No. 6.6;  
Matches 16; Conservative 3; Mismatches 13; Indels 30; Gaps 2;

OY 11 VARVRGGEGTYFVDFSV-----RNCPRH-----HFPHPN 40  
Db 74 VSCAASSGGYFVDFLRIGRPPQSLILADTGSDLVWKCACRNCSHSPATYVFFPHSS 133

OY 41 VF 42  
Db 134 TF 135

RESULT 15  
Q26791 PRELIMINARY: PRT; 1765 AA.  
ID Q26791:  
AC Q26791:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RNA POLYMERASE IIA LARGEST SUBUNIT (RPII197A).  
OS Trypanosoma brucei.  
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5691;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE-89168422; PubMed-2924350;  
RX Smith J.L., Levine J.R., Ingles C.J., Agabian N.;  
RT "In trypanosomes the homology of the largest subunit of RNA polymerase  
RT II is encoded by two genes and has a highly unusual C-terminal domain  
RT structure.";  
RL Cell 56:815-827(1989).  
DR EMBL: J03157; AAA30229.1; -  
DR InterPro: IPR002064; DNA\_pol\_B.  
DR InterPro: IPR000722; RNA\_pol\_A.  
DR InterPro: IPR002879; RNA\_pol\_A2.  
DR InterPro: IPR000822; znf-C2H2.  
DR Pfam: PF00623; RNA\_pol\_A; 1.  
DR Pfam: PF01854; RNA\_pol\_A2; 1.  
DR PROSITE: PS00116; DNA\_POLYMERASE\_B; UNKNOWN\_1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
SQ SEQUENCE 1765 AA; 196685 MW; ABIEICD05E283D25 CRC64;

Query Match 19.0%; Score 60.5; DB 5; Length 1765;  
Best Local Similarity 33.9%; Pred. No. 35;  
Matches 19; Conservative 7; Mismatches 23; Indels 7; Gaps 3;

OY 9 ERVAVRGG-----EGTYFVDFSVNCRPHRPNVFGCR-ADLFYDVEALDL 58  
Db 45 ERGRPVAGGINDLRMGTTDFEFACETCRKH-PECPGHFGYIELAEVFNIGVFDL 99

Search completed: May 24, 2002, 17:02:49  
Job time: 879 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:03:25 ; Search time 31.45 Seconds

(without alignments)  
71.406 Million cell updates/sec

Title: US-09-730-379E-2

Sequence: 1 ASFRVDRIERVARVRNGEGT.....PNVFGCRADLFYDEVALDL 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	307.5	96.7	525	1 HRG_HUMAN	P04196 homo sapien
2	219	68.9	526	1 HRG_RABIT	Q28640 oryctolagus
3	201	63.2	396	1 HRG_BOVIN	P33433 bos taurus
4	62	19.5	184	1 SODE_SCHMA	P16026 schistosoma
5	59.5	18.7	1403	1 VCG22_HSV11	000105 ictaluriid h
6	59	18.6	281	1 VNS1_INBST	P12600 influenza b
7	59	18.6	340	1 VNS1_INBPA	P12598 influenza b
8	58	18.2	235	1 TFP2_HUMAN	P48307 homo sapien
9	57.5	18.1	396	1 AAT_SALTI	O56114 salmonella
10	57.5	18.1	396	1 AAT_SALTY	P58661 salmonella
11	56.5	17.8	346	1 NRI1_ARATH	P32961 arabidopsis
12	55	17.3	516	1 YVNF_AZOCH	P24423 azotobacter
13	54.5	17.1	443	1 HEMI_HELMO	Q92996 heliobacill
14	54	17.0	329	1 VANA_PSES9	P12609 pseudomonas
15	54	17.0	452	1 ILK2_HUMAN	P57043 homo sapien
16	54	17.0	452	1 ILK_MOUSE	O55222 mus musculu
17	54	17.0	905	1 ZO3_MOUSE	O99xy1 mus musculu
18	54	17.0	1517	1 RPOC_CAMJE	O9p130 campylobact
19	54	17.0	2208	1 POLN_MANCV	Q69014 manchester
20	53.5	16.8	230	1 RL4_MYCLE	O32962 mycobacteri
21	53.5	16.8	466	1 VPI9_HSV2H	P22486 herpes simp
22	53.5	16.8	466	1 VPI9_HSV2H	P22486 herpes simp
23	53	16.7	1057	1 VPI2_AHSV3	O89508 african hor
24	52	16.4	571	1 PRDE_HUMAN	O99z88 homo sapien
25	52	16.4	676	1 NTP2_VACCA	O57193 vaccinia vi
26	52	16.4	676	1 NTP2_VACCC	P20502 vaccinia vi
27	52	16.4	676	1 NTP2_VACCT	O91fe3 vaccinia vi
28	52	16.4	676	1 NTP2_VACCV	P12927 vaccinia vi
29	52	16.4	676	1 NTP2_VARV	P33051 variola vir
30	52	16.4	933	1 ZO3_HUMAN	O95049 homo sapien
31	52	16.4	958	1 YGXI_YEAST	P53076 saccharomyc
32	51.5	16.2	213	1 PIS_HUMAN	O14735 homo sapien
33	51.5	16.2	213	1 PIS_RAT	P70500 rattus norv

34	51.5	16.2	231	1 DLH1_AOUAE	O67802 aquifex aeo
35	51.5	16.2	346	1 NRI3_ARATH	P46010 arabidopsis
36	51.5	16.2	711	1 LCPI_MOUSE	O99u17 mus musculu
37	51	16.0	171	1 CFIA_HSYE2	O66674 equine herp
38	51	16.0	281	1 VNS1_INBHT	P12595 influenza b
39	51	16.0	451	1 ILK_CAVPO	P57044 cavla porce
40	51	16.0	452	1 ILK1_HUMAN	O13418 homo sapien
41	51	16.0	862	1 LOXA_PHAUV	P27480 phaseolus v
42	51	16.0	863	1 AD17_DROME	O99ae5 drosophila
43	51	16.0	1434	1 NOST_HUMAN	P29475 homo sapien
44	50.5	15.9	93	1 AAT_METEX	P52069 methylolact
45	50.5	15.9	165	1 YB81_YEAST	P38362 saccharomyc

## ALIGNMENTS

```

RESULT 1
ID      HRG_HUMAN      STANDARD;      PRT;      525 AA.
AC      P04196;
DT      20-MAR-1987 (Rel. 04, Created)
DT      20-MAR-1987 (Rel. 04, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Histidine-rich glycoprotein precursor (Histidine-proline rich
DE      glycoprotein) (HPRG).
GN      HRG.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=66216149; PubMed=3011081;
RA      Koide T., Foster D.C., Yoshitake S., Davie E.W.;
RT      "Amino acid sequence of human histidine-rich glycoprotein derived
RT      from the nucleotide sequence of its cDNA.";
RL      Biochemistry 25:2220-2225(1986).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;
RL      Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 214-247 FROM N.A.
RX      MEDLINE=94245171; PubMed=8188234;
RA      Henmans B.C., Frans R.R., Bakker E., Vossen R.H., van der Poort E.W.,
RT      Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;
RT      "Evidence for the absence of intron H of the histidine-rich
RT      glycoprotein (HRG) gene: genetic mapping and in situ localization of
RT      HRG to chromosome 3q28-q29.";
RL      Genomics 19:195-197(1994).
RN      [4]
RP      SEQUENCE OF 19-27.
RX      TISSUE-Plasma;
RA      Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA      Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA      Hochstrasser D.F.;
RT      "Plasma protein map: an update by microsequencing.";
RL      Electrophoresis 13:707-714(1992).
RN      [5]
RP      FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
RN      [6]
RP      HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
RN      [7]
RP      FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
RN      [8]
RP      AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS
RN      [9]
RP      HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
RN      [10]
RP      MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
RN      [11]
RP      COAGULATION CASCADE.
RN      [12]
RP      -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
RN      [13]
RP      PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATITIONS OF A 5-
RN      [14]
RP      RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.
RN      [15]
RP      -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
RN      [16]
RP      -----
RN      [17]
RP      THIS SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M13149; AAA52694.1; -  
DR EMBL; AB005803; BAA21613.1; -  
DR EMBL; 217218; CAA78925.1; -  
DR PIR; A01287; KGHUGH.  
DR SWISS-2DPAGE; P04196; HUMAN.  
DR MIM; 142640; -  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF000031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
KM Glycoprotein; Heparin-binding; Repeat; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
FT DOMAIN 19 136 CYSTATIN-LIKE 1.  
FT DOMAIN 137 254 CYSTATIN-LIKE 2.  
FT DOMAIN 276 321 PRO-RICH.  
FT DOMAIN 350 497 PRO/HIS-RICH.  
FT DISULFID 24 504 BY SIMILARITY.  
FT DISULFID 78 89 BY SIMILARITY.  
FT DISULFID 105 126 BY SIMILARITY.  
FT DISULFID 203 417 BY SIMILARITY.  
FT DISULFID 218 241 BY SIMILARITY.  
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 96.7%; Score 307.5; DB 1; Length 525;  
Best Local Similarity 98.3%; Pred. No. 5.5e-33;  
Matches 58; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ASFRVDRIEVAVRGEGT-YFVDFSVRNCPRHHPHNVFGFCRADLFYDEALD 58  
DB 173 ASFRVDRIEVAVRGEGTGYFDFSVRNCPRHHPHNVFGFCRADLFYDEALD 231

RESULT 2  
HRG\_RABIT STANDARD: PRT; 526 AA.  
AC Q28640;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
DE glycoprotein) (HPRG) (Fragment).  
GN HRG.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
RC TISSUE=Serum;  
RX MEDLINE=96229917; PubMed=8639676;  
RA Borza D.-B., Tatum F.M., Morgan W.T.;  
RT "Domain structure and conformation of histidine-proline-rich  
RT glycoprotein".  
RL Biochemistry 35:1925-1934(1996).  
-1- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD

CC COAGULATION CASCADE.  
CC -1- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
CC PROTEIN HAS MANY INTERNAL REPEATS. 15 TANDEM REPEATITIONS OF A 5-  
CC RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH  
CC REGION.  
CC -1- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U32189; AAC48516.1; -  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF000031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
KM Glycoprotein; Heparin-binding; Repeat; Signal.  
FT SIGNAL 1 8  
FT NON\_TER 1  
FT CHAIN 9 526  
FT DOMAIN 9 126 HISTIDINE-RICH GLYCOPROTEIN.  
FT DOMAIN 127 243 CYSTATIN-LIKE 1.  
FT DOMAIN 251 296 CYSTATIN-LIKE 2.  
FT DOMAIN 329 498 PRO-RICH.  
FT DISULFID 14 505 PRO/HIS-RICH.  
FT DISULFID 68 79 BY SIMILARITY.  
FT DISULFID 95 116 BY SIMILARITY.  
FT DISULFID 193 415 BY SIMILARITY.  
FT DISULFID 207 230 BY SIMILARITY.  
FT DISULFID 272 302 POTENTIAL.  
FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 303 304 CLEAVAGE (BY PLASMIN).  
FT SITE 421 422 CLEAVAGE (BY PLASMIN).  
SQ SEQUENCE 526 AA; 58877 MW; 810F23D367D93D42 CRC64;

Query Match 68.9%; Score 219; DB 1; Length 526;  
Best Local Similarity 74.6%; Pred. No. 2.6e-21;  
Matches 44; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

QY 1 ASFRVDRIEVAVRGEGT-YFVDFSVRNCPRHHPHNVFGFCRADLFYDEALD 58  
DB 163 ASFRVDRIEVAVRGEGTGYFDFSVRNCPRHHPHNVFGFCRADLFYDEASNL 220

RESULT 3  
HRG\_BOVIN STANDARD: PRT; 396 AA.  
AC P33433;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)  
DE (HPRG) (Fragments).  
GN HRG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE, AND DISULFIDE BONDS.  
RX MEDLINE=93351678; PubMed=8348977;  
RA Soerensen C.B., Krogh-Pedersen H., Petersen T.E.;  
RT "determination of the disulphide bridge arrangement of bovine  
RT histidine-rich glycoprotein".



```
CC -I- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -I- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -I- COFACTOR: Copper and zinc.  
CC -I- SUBUNIT: HOMODIMER.  
CC -I- SIMILARITY LOCATION: Extracellular.  
CC -I- SIMILARITY BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@lsb.sib.ch).  
-----  
DR EMBL; M27529; AAA29937.1; -.  
DR EMBL; M28545; AAA29934.1; -.  
DR EMBL; M28543; AAA29934.1; JOINED.  
DR EMBL; M28544; AAA29934.1; JOINED.  
DR PIR; A37019; A37019.  
DR HSPP; P07505; ISRD.  
DR InterPro; IPR001424; SOD_CU_ZN.  
DR Pfam; PF00068; sodcu: 1.  
DR PRINTS; PR00068; CUZNDISMUTASE.  
DR PRODOM; PD000469; SOD_CU_ZN_1;  
DR PROSITE; PS00087; SOD_CU_ZN_1;  
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.  
KW Oxidoreductase; Copper; zinc; signal.  
  
FT SIGNAL 1 18  
FT CHAIN 19 184 EXTRACELLULAR SUPEROXIDE DISMUTASE [CU-  
FT METAL ZN].  
FT METAL 76 76 COPPER (BY SIMILARITY).  
FT METAL 78 78 COPPER (BY SIMILARITY).  
FT METAL 93 93 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 101 101 ZINC (BY SIMILARITY).  
FT METAL 110 110 ZINC (BY SIMILARITY).  
FT METAL 113 113 ZINC (BY SIMILARITY).  
FT METAL 150 150 COPPER (BY SIMILARITY).  
FT DISULFID 87 176 BY SIMILARITY.  
FT CARBOHYD 61 61 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 184 AA; 20346 MW; C14FD37026FA8BB3 CNC64;
```

Query Match 19.5%; Score 62; DB 1; Length 184;  
Best Local Similarity 30.4%; Pred.No. 0.46;  
Matches 14; Conservative 7; Mismatches 17; Indels 8; Gaps 1.

Oy 19 GTTYEVDSEVRNCPRHHFRHPNVFGFCFA-----DLFFDEVEL 56  
| : | : | : | : ||| | : | : | : |  
Db 91 GRHFPEFNQRHGGRPHRGYPRHAGDLLGNIVRGGVAKPDFYVTIKGL 136

RESULT 5  
VG22\_HSV11 STANDARD; PROT; 1403 AA.  
AC Q00105;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DE Hypothetical gene 22 protein.  
GN 22.  
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC unclassified Herpesviridae.  
OX NCBI\_TaxId=10401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=AUBURN 1;  
RC MEDLINE=92087490; PubMed=1727613;  
RA Davison A.J.;  
RT "Channel catfish virus: a new type of herpesvirus.";  
RL Virology 186:9-14(1992).

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M75136; AAA88125.1; -
DR PIR: E36788; E36788.
DR HSSP: P05766; 1A32.
KM Hypothetical protein.
SQ SEQUENCE 1403 AA; 153308 MW; FAGE366FC04A33AE CRC64;

Query Match 18.7%; Score 59.5; DB 1; Length 1403;
Best Local Similarity 30.4%; Pred. No. 8.4;
Matches 17; Conservative 9; Mismatches 27; Indels 3; Gaps 1;

QY 2 SFVRDRIERVARVGGEGTYFVDSVNCPRHHPRRHNVFGCRADLFYDVEALD 57
Db 71 SVRDVQRIEEL---GTVSDSDSVEFVLEAHLSKVKDIFGNSVTWADYKAD 123

RESULT 6
VNS1_INBSI STANDARD; PRT; 281 AA.
AC P12600;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nonstructural protein NS1.
OS Influenza B virus (strain B/Singapore/222/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza B virus.
OX NCBI_TaxID=11544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86160037; Pubmed=3267218;
RA Yamashita M., Krystal M., Fitch W.M., Palese P.;
RT "Influenza B virus evolution: co-circulating lineages and comparison
RT of evolutionary pattern with those of influenza A and C viruses.";
RL Virology 163:112-122(1988).
CC -1- ALTERNATIVE PRODUCTS: NS1 AND NS2 PROTEINS ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF SEGMENT 8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19790; AAA43775.1; -
DR InterPro: IPR004208; FLU_B_NS1.
DR Pfam: PF02942; FLU_B_NS1; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 281 AA; 32156 MW; ADFAF70C6C4820A3 CRC64;

Query Match 18.6%; Score 59; DB 1; Length 281;
Best Local Similarity 30.4%; Pred. No. 1.8;
Matches 14; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 12 ARVAGCGGTFFVDSVNCPRHHPRRHNVFGCRADLFYDVEALD 57
Db 95 AGIEGFPCMKNFNSNCNPYMTDYPPTPGKCLDIEEPENVD 140

RESULT 7
VNS1_INBSI STANDARD; PRT; 281 AA.

```

```

ID VNS1_INBSI STANDARD; PRT; 340 AA.
AC P12598;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nonstructural protein NS1.
OS Influenza B virus (strain B/PA/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza B virus.
OX NCBI_TaxID=11542;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86160037; Pubmed=3267218;
RA Yamashita M., Krystal M., Fitch W.M., Palese P.;
RT "Influenza B virus evolution: co-circulating lineages and comparison
RT of evolutionary pattern with those of influenza A and C viruses.";
RL Virology 163:112-122(1988).
CC -1- ALTERNATIVE PRODUCTS: NS1 AND NS2 PROTEINS ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF SEGMENT 8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19791; AAA43764.1; -
DR InterPro: IPR004208; FLU_B_NS1.
DR InterPro: IPR000968; FLU_NS2.
DR Pfam: PF02942; FLU_B_NS1; 1.
DR Pfam: PF00601; FLU_NS2; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 340 AA; 39541 MW; E79DE2883272B5A8 CRC64;

Query Match 18.6%; Score 59; DB 1; Length 340;
Best Local Similarity 30.4%; Pred. No. 2.2;
Matches 14; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 12 ARVAGCGGTFFVDSVNCPRHHPRRHNVFGCRADLFYDVEALD 57
Db 95 AGIEGFPCMKNFNSNCNPYMTDYPPTPGKCLDIEEPENVD 140

RESULT 8
TFP2_HUMAN STANDARD; PRT; 235 AA.
ID TFP2_HUMAN
AC P48307;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tissue factor pathway inhibitor 2 precursor (TFPI-2) (Placental
DE protein 5) (PP5).
GN TFP12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=95204397; Pubmed=7896752;
RA Miyagi Y., Koshikawa N., Yasumitsu H., Miyagi E., Hirahara F.,
RA Aoki I., Misugi K., Umeda M., Miyazaki K.;
RT "cDNA cloning and mRNA expression of a serine proteinase inhibitor
RT secreted by cancer cells: identification as placental protein 5 and
RT tissue factor pathway inhibitor-2.";
RL J. Biochem. 116:939-942(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;

```

[illegible]

KW Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.  
 FT BINDING 246 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT ACT\_SITE 374 BINDS THE SUBSTRATE ALPHA-CARBOXYLATE (BY  
 FT SIMILARITY).  
 SQ SEQUENCE 396 AA; 43507 MW; 974C1585438D02FB CRC64;

Query Match 18.1%; Score 57.5; DB 1; Length 396;  
 Best Local Similarity 29.0%; Pred. No. 4.1;  
 Matches 18; Conservative 7; Mismatches 24; Indels 13; Gaps 3;

QY 6 DRIERVARVAGGEGTVFV--DFSVRNC-----RHFFRHPVVEGFC----RADLFYD 52  
 ID NRRL\_ARATH STANDARD; PRT; 346 AA.  
 AC P32961; Q42543; 004908;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 1 (EC 3.5.5.11)  
 GN NIT1 OR AT3644310 OR T10D17\_100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

QY 53 VE 54  
 Db 152 AE 153

RESULT 10

AAR\_SALTY ID AAT\_SALTY STANDARD; PRT; 396 AA.  
 AC P58661;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Aspartate aminotransferase (EC 2.6.1.1) (Transaminase A) (Aspat).  
 GN ASPC OR STMO099.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lettelle P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RA \*Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.  
 RL Nature 413:852-856(2001).  
 CC -1- CATALYTIC ACTIVITY: L-aspartate + 2-oxoglutarate = oxaloacetate +  
 CC L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE008743; AAL19932.1; -;  
 DR StyGene; SG77777; aspc.  
 KW Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.  
 FT BINDING 246 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT ACT\_SITE 374 BINDS THE SUBSTRATE ALPHA-CARBOXYLATE (BY  
 FT SIMILARITY).  
 SQ SEQUENCE 396 AA; 43521 MW; 96094D0D148D02FB CRC64;

Query Match 18.1%; Score 57.5; DB 1; Length 396;  
 Best Local Similarity 29.0%; Pred. No. 4.1;  
 Matches 18; Conservative 7; Mismatches 24; Indels 13; Gaps 3;

QY 6 DRIERVARVAGGEGTVFV--DFSVRNC-----RHFFRHPVVEGFC----RADLFYD 52  
 ID NRRL\_ARATH STANDARD; PRT; 346 AA.  
 AC P32961; Q42543; 004908;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 1 (EC 3.5.5.11)  
 GN NIT1 OR AT3644310 OR T10D17\_100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

QY 53 VE 54  
 Db 152 AE 153

RESULT 11

ID NRRL\_ARATH STANDARD; PRT; 346 AA.  
 AC P32961; Q42543; 004908;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 1 (EC 3.5.5.11)  
 GN NIT1 OR AT3644310 OR T10D17\_100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=CV. LANDSBERG ERRECTA; TISSUE=Leaf.  
 RX MEDLINE=92209532; PubMed=1555601;  
 RA Bartling D., Seedorf M., Mitchofer A., Weller E.W.;  
 RT "Cloning and expression of an Arabidopsis nitrilase which can convert  
 RT indole-3-acetonitrile to the plant hormone, indole-3-acetic acid.";  
 RL Eur. J. Biochem. 205:417-424(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Zhou L., Bartel B., Thornburg R.W.;  
 RT "Nucleotide sequence of the Arabidopsis thaliana nitrilase 1 gene.";  
 RL (in) Plant Gene Register PGR95-130.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98145459; PubMed=9484465;  
 RA Hillebrand H., Bartling D., Weller E.W.;  
 RT "Structural analysis of the nitr2/nit1/nit3 gene cluster encoding  
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-  
 RT acetic acid biosynthesis in Arabidopsis thaliana.";  
 RL Plant Mol. Biol. 36:89-99(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unselid M.,  
 RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Delenly M., Boutry M., Griwell L.A., Mache R., Puigdomenech P.,  
 RA de Simone V., Choise N., Artiguenave F., Robert C., Brotier P.,  
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Queller F.,  
 RA Schaefer M., Weller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Reichelt J., Schafte M., Schoen O., Bargues M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwalder B., Duchemin D.,  
 RA Cooke R., Laude M., Berger-Llauró C., Purnelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Montfort A., Argirou A., Flores M., Liguori R., Vitale D.,  
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Watts A., Utterback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,  
 RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,

```

RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Matsumoto A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
RN [5]
RP CHARACTERIZATION.
RC STRAIN=CV. LANDSBERG. ERECTA;
RX MEDLINE=94286570; PubMed=8016109;
RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOLE-3-ACETIC ACID.
CC -1- CATALYTIC ACTIVITY: A nitrile + H(2)O = a carboxylate + NH(3).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A
CC VERY LOW LEVEL DURING THE FRUITING STAGE.
CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X63445; CAA5041.1; -
DR EMBL: U38845; AAB05221.1; -
DR EMBL: Y07648; CAA68935.2; -
DR EMBL: AL353865; CAB88999.1; -
DR PIR: S22398; S22398.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
KW Hydrolase; Multigene family.
FT ACCT_SITE 186
FT CONFLICT 312 312 Y->H (IN REF. 2).
SO SEQUENCE 346 AA; 38178 MW; 8DAF87CAD1E3C1F CRC64;

Query Match 17.8%; Score 56.5; DB 1; Length 346;
Best Local Similarity 37.8%; Pred. No. 4.8;
Matches 14; Conservative 5; Mismatches 15; Indels 3; Gaps 1;

QY 18 EGYTFVDFSVRNCGRHNRHNVFGRADLFYDVE 54
DB 227 EGGCFVLACQFCQKRNHRPDHPD--YLTPTWYDKX 260

RESULT 12
YVNF_AZCH STANDARD; PRT; 516 AA.
AC P24423;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in vnfD 5' region.
OS Azotobacter chroococcum mcd 1.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=355;
RN [1]
RP MEDLINE=90356423; PubMed=2388847;
RA Fallik E., Robson R.L.;
RT "Completed sequence of the region encoding the structural genes for
RT the vanadium nitrogenase of Azotobacter chroococcum.";
```

```

RL Nucleic Acids Res. 18:4616-4616(1990).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X51756; CAA36057.1; -
DR PIR: S14694; S14694.
KW Nitrogen fixation; Hypothetical protein.
SO SEQUENCE 516 AA; 58676 MW; ED7A7F362ED08A14 CRC64;

Query Match 17.3%; Score 55; DB 1; Length 516;
Best Local Similarity 32.1%; Pred. No. 11;
Matches 17; Conservative 10; Mismatches 18; Indels 8; Gaps 3;

QY 8 IERVARRGEGTYFVD--FSVRNCR--HHPRHNVFGRADLFYDVEA 55
DB 10 IQFFAKIREGDYDYVDKTSFALQLIQGTHYFLSRPRRCG--KSLFDTLA 59

RESULT 13
HEML_HELMO STANDARD; PRT; 443 AA.
ID HEML_HELMO
AC Q9ZG56;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.1-) (GluTR).
GN HEMA.
OS Helibacillus mobilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Helibacterium group; Helibacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
RT a major photosynthesis gene cluster from Helibacillus mobilis-".
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA(GLU) + NADPH = GLUTAMATE-1-
CC SEMIALDEHYDE + NADP(+) + TRNA(GLU).
CC -1- PATHWAY: FIRST STEP IN PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF080002; AAC84013.1; -
DR HSSP: Q42843; 1B61.
DR InterPro: IPR000343; GluTR.
DR Pfam: PF00745; GluTR; 1.
DR PROSITE: PS00747; GluTR; 1.
KW Porphyrin biosynthesis; Oxidoreductase; NADP.
SO SEQUENCE 443 AA; 49577 MW; E95C30E251A0C5F6 CRC64;

Query Match 17.1%; Score 54.5; DB 1; Length 443;
Best Local Similarity 32.1%; Pred. No. 11;
Matches 17; Conservative 8; Mismatches 21; Indels 7; Gaps 2;

QY 4 RVDIERVARRGEGTYFVDFSVRNCGRHNRHNVFGRADLFYDVEAL 56
DB 10 IQFFAKIREGDYDYVDKTSFALQLIQGTHYFLSRPRRCG--KSLFDTLA 59
```

RESULT	14
ID	VANA_PSES9
STANDARD:	PRT: 329 AA.
AC	P12609;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-OCT-1989 (Rel. 12, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Vanillate O-demethylase oxygenase subunit (EC 1.2.3.12) (4-hydroxy-3-methoxybenzoate demethylase).
GN	VANA.
OS	Pseudomonas sp. (strain ATCC 19151).
OC	Bacteria; Proteobacteria.
NCBI_TaxID=315;	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89008117; PubMed=3170489;
RA	Brunel F., Davison J.;
RT	"Cloning and sequencing of Pseudomonas genes encoding vanillate demethylase".;
RL	J. Bacteriol. 170:4924-4930(1988).
CC	-1- CATALYTIC ACTIVITY: Vanillate + O(2) + NADH = 3,4-dihydroxybenzoate + NAD(+) + H(2)O + formaldehyde.
CC	-1- COPACITOR: PROBABLY BINDS A 2FE-2S GROUP AND AN IRON ATOM.
CC	-1- PATHWAY: VANILLATE DEGRADATION (VANILLATE IS A KEY INTERMEDIATE IN THE DEGRADATION OF LIGNIN).
CC	-1- SUBUNIT: THIS DEMETHYLASE SYSTEM CONSISTS OF TWO PROTEINS: AN OXYGENASE AND AN OXYGENASE REDUCTASE.
CC	-1- SIMILARITY: BELONGS TO THE BACTERIAL RING-HYDROXYLATING DIOXYGENASE ALPHA SUBUNIT FAMILY.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .
CC	-----
DR	EMBL; M22077; AAA26019.1; .
DR	PIR; A43652; A43652.
DR	InterPro: IPR001281; Rleske.
DR	InterPro: IPR001663; Ring_hydroxyl_A.
DR	Pfam; PF00355; Rleske; 1.
DR	PROSITE; PS00570; RING-HYDROXYL_ALPHA; FALSE_NEG.
KW	Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
FT	Iron-sulfur; Iron; NAD; Lignin degradation.
FT	METAL 24 24 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT	METAL 26 26 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT	METAL 43 43 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT	METAL 46 46 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SO	SEQUENCE 329 AA; 36578 MW; BAEABD469E92BF43 CRC64;
QY	2 SFRVDRIRVARVRGEGSTYFVD 24
Db	252 SFRPEDNELTARIRRGOGTIFAE 274
Query Match	17.0%; Score 54; DB 1; Length 329;
Best Local Similarity	47.8%; Pred. No. 9.7;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;	
RESULT 15	
ID	ILK2_HUMAN
STANDARD:	PRT: 452 AA.
AC	P57043;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Integrin-linked protein kinase 2 (EC 2.7.1.-) (ILK-2).

Query Match	17.0%	Score 54;	DB 1;	Length 452;
Best Local Similarity	35.6%	Pred. No. 14;		
Matches 16;	Conservative 5;	Mismatches 20;	Indels 4;	Gaps 2;



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:50:09 ; Search time 64.04 Seconds  
(without alignments)  
87.027 Million cell updates/sec

Title: US-09-730-379E-2

Perfect score: 318  
Sequence: 1 ASFRVDRIEFVARVARGEGCT.....PNVFGFCRADLFYDEALDL 58

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	307.5	96.7	525	1 KGHUGH	histidine-rich gly
2	212	66.7	445	2 A60488	histidine-rich gly
3	62	19.5	184	2 A37019	superoxide dismuta
4	60.5	19.0	1765	2 A31494	DNA-directed RNA p
5	60.5	19.0	1765	2 B31494	DNA-directed RNA p
6	60	18.9	855	2 S77252	phosphorylase (EC
7	59.5	18.7	1404	2 E36788	hypothetical prote
8	59	18.6	218	2 S75100	ABC transport prot
9	58.5	18.4	326	2 G69366	homoserine dehydro
10	58	18.2	235	2 A54951	tissue factor path
11	57.5	18.1	396	2 AD0616	aspartate aminotra
12	56.5	17.8	346	2 T49147	nitrilase (EC 3.5.
13	56.5	17.8	346	2 S23398	nitrilase (EC 3.5.
14	56	17.6	1350	2 T10803	probable RNA direc
15	55.5	17.5	132	2 S75953	hypothetical prote
16	55.5	17.5	166	2 S15162	probable flavin mo
17	55.5	17.5	527	2 B70920	hypothetical prote
18	55.5	17.5	1580	2 T26204	hypothetical prote
19	55	17.3	446	2 D71418	hypothetical prote
20	55	17.3	516	2 S14694	hypothetical prote
21	54.5	17.1	306	2 G96935	mecf-like protein
22	54.5	17.1	443	2 T31441	probable glutamyl-
23	54.5	17.1	571	2 F70040	sulfite reductase
24	54.5	17.1	1559	2 T07757	probable DNA (cyto
25	54	17.0	140	2 T14771	hypothetical prote
26	54	17.0	309	2 T33376	hypothetical prote
27	54	17.0	310	2 T51694	D-alanine-D-alani
28	54	17.0	329	2 A43652	probable monooxyge
29	54	17.0	391	2 H84298	2',3'-cyclic-nucle

30	54	17.0	1517	2 B81393	DNA-directed RNA p
31	53.5	16.8	230	2 T45365	ribosomal protein
32	53.5	16.8	252	2 T46661	beta-1,4-glucosylt
33	53.5	16.8	252	2 B81053	beta-1,4-glucosylt
34	53.5	16.8	254	2 D81824	beta-1,4-glucosylt
35	53.5	16.8	466	1 WMBEHT	capsid protein ICP
36	53.5	16.8	546	2 JC4798	seizure-related me
37	53.5	16.8	1469	2 T19459	hypothetical prote
38	53	16.7	221	2 AB2204	lipoate-protein 11
39	53	16.7	1057	2 S47151	outer capsid prote
40	52.5	16.5	396	2 C90755	aspartate aminotra
41	52.5	16.5	396	2 A85619	aspartate aminotra
42	52.5	16.5	683	2 B71325	conserved hypothet
43	52	16.4	152	2 AD2512	hypothetical prote
44	52	16.4	235	2 AF2319	hypothetical prote
45	52	16.4	495	2 AD1927	hypothetical prote

## ALIGNMENTS

RESULT 1  
KGHUGH  
histidine-rich glycoprotein precursor - human  
N:Alternate names: HRG  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000  
C:Accession: A01287; S29669  
R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.  
Biochemistry 25, 2220-2225, 1986  
A>Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu  
A:Reference number: A01287; MUID:86216149  
A:Accession: A01287  
A:Molecule type: mRNA  
A:Residues: 1-525 <KOI>  
A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BA21613.1; PID:g2280514  
R:Hennis, B.; Havelaar, A.; Kuft, C.  
submitted to the EMBL Data Library, October 1991  
A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly  
A:Reference number: S29669  
A:Accession: S29669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 214-247 <HEN>  
A:Cross-references: EMBL:U17218; NID:g32453; PIDN:CA78925.1; PID:g32454  
C:Comment: Although its physiological function is not yet known, HRG does bind heme,  
din, and the lysine-binding site of plasminogen. On the basis of its homology with Hm  
lood coagulation cascade.  
C:Comment: The amino half of this protein is homologous to the first two cystatin-11k  
could not have inhibitory activity.  
C:Comment: In addition to having a high histidine and proline content, this protein h  
e-rich' region.  
C:Genetics:  
A:Gene: GDB:HRG  
A:Cross-references: GDB:120055; OMIM:142640  
A:Map position: 3q27-3q27  
C:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat  
F:1-16/Domain: signal sequence #status predicted <Sto>  
F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>  
F:19-131/Domain: cystatin homology <CY1>  
F:140-246/Domain: cystatin homology <CY2>  
F:276-321/Region: proline-rich  
F:348-437/Region: histidine-rich  
F:551-497/Region: proline-rich  
F:63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 96.7%; Score 307.5; DB 1; Length 525;  
Best Local Similarity 98.3%; Pred. No. 1,6e-31;  
Matches 58; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 ASFVRDRIERYARVGGEGT-YFVDFSVRNCPRHHFPRHNVGFCRADLFYDVEALDL 58  
 Db 173 ASFVRDRIERYARVGGEGTGYFDFSVRNCPRHHFPRHNVGFCRADLFYDVEALDL 231

## RESULT 2

A60488  
 histidine-rich glycoprotein - bovine (fragments)  
 N:Alternate names: autorosette inhibition factor  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 19-Mar-1993 #sequence\_revision 23-Mar-1995 #text\_change 07-Jul-1995  
 C:Accession: S35687; J02196; A60488  
 R:Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.  
 FEBS Lett. 328, 285-290, 1993  
 A:Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly  
 A:Reference number: S35687; MUID:93351678  
 A:Accession: S35687  
 A:Molecule type: protein  
 A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>  
 A:Note: 355-Gln and 368-Tyr were also found  
 R:Haliker, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.  
 Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
 A:Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
 A:Reference number: J02196; MUID:94220160  
 A:Accession: J02196  
 A:Molecule type: protein  
 A:Residues: 1-23;35-54; 'VK', 57-101, 'R', 'TVGEYEG', 116, 'N', 118, 'R', 120-136;137-145;150-20  
 A:Experimental source: Plasma  
 R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Haliker, T.  
 Thromb. Res. 60, 383-396, 1990  
 A:Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
 A:Reference number: A60488; MUID:91196010  
 A:Accession: A60488  
 A:Molecule type: protein  
 A:Residues: 1-6, 'X', 8-15 <YES>  
 C:Comment: This protein is a single-chained plasma protein which participates in transgl  
 C:Superfamily: histidine-rich glycoprotein; cystatin homology  
 C:Keywords: glycoprotein; plasma  
 F:2-113/Domain: cystatin homology <CY1>  
 F:122-207/Domain: cystatin homology (fragments) <CY2>  
 F:7-424,60-71,87-108,165-346,180-203,258-288/Disulfide bonds: #status experimental  
 F:74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 66.7%; Score 212; DB 2; Length 445;  
 Best Local Similarity 69.5%; Pred. No. 2,5e-19;  
 Matches 41; Conservative 4; Mismatches 6; Indels 8; Gaps 2;

OY 1 ASFVRDRIERYARVGGEGT-YFVDFSVRNCPRHHFPRHNVGFCRADLFYDVEALDL 58  
 Db 142 AFRPM-----RARGEGTSYFLDFSVRNCSSHHFPRSHIFGFCRADLFYDVEASDL 193

## RESULT 3

A37019  
 superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - fluke (Schistosoma mansoni)  
 C:Species: Schistosoma mansoni  
 C>Date: 24-Jan-1992 #sequence\_revision 24-Jan-1992 #text\_change 20-Apr-2000  
 C:Accession: A37019  
 R:Simurda, M.C.; van Keulen, H.; Rekosh, D.M.; Loverde, P.T.  
 Exp. Parasitol. 67, 73-84, 1988  
 A:Title: Schistosoma mansoni: identification and analysis of an mRNA and a gene encoding  
 A:Reference number: A37019; MUID:89005535  
 A:Accession: A37019  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-184 <STM>  
 A:Cross-references: GB:M28543; GB:M28544; GB:M28545; NID:g161115; PIDN:AAA29934.1; PID:g  
 C:Function:  
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
 C:Superfamily: superoxide dismutase (Cu-Zn)  
 C:Keywords: copper; metalloprotein; oxidoreductase; zinc  
 F:173/Active site: Arg #status predicted

Query Match 19.5%; Score 62; DB 2; Length 184;  
 Best Local Similarity 30.4%; Pred. No. 1.9;  
 Matches 14; Conservative 7; Mismatches 17; Indels 8; Gaps 1;

OY 19 GTYFVDFSVRNCPRHHFPRHNVGFCRA-----DLFYDVEAL 56  
 Db 91 GPHFNPQNRQHRGHPYRAGDGLNIRVGRGVAKDFVYTIKGL 136

## RESULT 4

A31494  
 DNA-directed RNA polymerase (EC 2.7.7.6) IIA - Trypanosoma brucei  
 C:Species: Trypanosoma brucei  
 C>Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 23-Apr-1999  
 C:Accession: A31494  
 R:Smith, J.L.; Levin, J.R.; Ingles, C.J.; Agabian, N.  
 Cell 56, 815-827, 1989  
 A:Title: In trypanosomes the homolog of the largest subunit of RNA polymerase II is e  
 A:Reference number: A31494; MUID:89168422  
 A:Accession: A31494  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1765 <SMI>  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: nucleotidyltransferase; transcription

Query Match 19.0%; Score 60.5; DB 2; Length 1765;  
 Best Local Similarity 33.9%; Pred. No. 27;  
 Matches 19; Conservative 7; Mismatches 23; Indels 7; Gaps 3;

OY 9 ERVARVAGG-----EGTYFVDFSVRNCPRHHFPRHNVGFCRA-ADLFYDVEALDL 58  
 Db 45 ERGRPVAGGINDLRMGTTDFEFACETCHRKH-FDCRPHFYIELAEVFNIGVFDL 99

## RESULT 5

B31494  
 DNA-directed RNA polymerase (EC 2.7.7.6) IIB - Trypanosoma brucei  
 C:Species: Trypanosoma brucei  
 C>Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 23-Apr-1999  
 C:Accession: B31494  
 R:Smith, J.L.; Levin, J.R.; Ingles, C.J.; Agabian, N.  
 Cell 56, 815-827, 1989  
 A:Title: In trypanosomes the homolog of the largest subunit of RNA polymerase II is e  
 A:Reference number: A31494; MUID:89168422  
 A:Accession: B31494  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1765 <SMI>  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: nucleotidyltransferase; transcription

Query Match 19.0%; Score 60.5; DB 2; Length 1765;  
 Best Local Similarity 33.9%; Pred. No. 27;  
 Matches 19; Conservative 7; Mismatches 23; Indels 7; Gaps 3;

OY 9 ERVARVAGG-----EGTYFVDFSVRNCPRHHFPRHNVGFCRA-ADLFYDVEALDL 58  
 Db 45 ERGRPVAGGINDLRMGTTDFEFACETCHRKH-FDCRPHFYIELAEVFNIGVFDL 99

## RESULT 6

S77252  
 phosphorylase (EC 2.4.1.1) 2 - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein S11367  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S77252

R.Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 S.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S77252  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-855 <KAN>  
 A:Cross-references: EMBL:D90907; GB:AB001339; NID:g1652618; PIDN:BAI17586.1; PID:g165266  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Keywords: ATP; nucleotide binding; P-loop  
 F:17-210/Domain: ATP-binding cassette homology <ABC>  
 F:34-41/Region: nucleotide-binding motif A (P-loop)  
 A:Gene: glp-2  
 A:Start codon: GTG  
 C:Superfamily: phosphorylase  
 C:Keywords: glycosyltransferase; hexosyltransferase; phosphoprotein; pyridoxal phosphate  
 F:691/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 18.9%; Score 60; DB 2; Length 855;  
 Best Local Similarity 34.2%; Pred. No. 15;  
 Matches 13; Conservative 8; Mismatches 11; Indels 6; Gaps 1;

OY 21 YEVDSVRCPCRRHPPHPPVFGFCRADLFYDVALDL 58  
 :|||:::| | |||: | ||::| |  
 Db 317 FFVSASLDLIRLRLHRLTHPML-----DFEYELTAIQL 348

RESULT 7  
 hypothetical protein ORF22 - ictalurid herpesvirus 1 (strain auburn 1)  
 C:Species: ictalurid herpesvirus 1  
 A:Note: host ictalurid punctatus (channel catfish)  
 C:Date: 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change 08-Oct-1999  
 C:Accession: E36788  
 R:Davidson, A.J.  
 A:Submitted to GenBank, January 1992  
 A:Description: Channel catfish virus: a new type of herpesvirus.  
 A:Reference number: A36804  
 A:Accession: E36788  
 A:Molecule type: DNA  
 A:Residues: 1-1404 <DAV>  
 A:Cross-references: GB:M75136; NID:g331209; PIDN:AAA88125.1; PID:g331232  
 R:Davidson, A.J.  
 A:Title: Channel catfish virus: a new type of herpesvirus.  
 A:Reference number: A39447; MUID:92087490  
 A:Contents: annotation  
 A:Note: neither protein nor nucleic acid sequence is given  
 C:Genetics:  
 A:Gene: 22

Query Match 18.7%; Score 59.5; DB 2; Length 1404;  
 Best Local Similarity 30.4%; Pred. No. 29;  
 Matches 17; Conservative 9; Mismatches 27; Indels 3; Gaps 1;

OY 2 SRRVNIIEVAVRGGEGTYFDVSVRNCPRHHFRRHPPVFGFCRADLFYDVALD 57  
 :|||:::| | |||: | ||::| |  
 Db 72 SVRMVLIQRIEEL--GTTVSDDSVFEVIEAHLKVKDIFGNCSTWTWADVKAAD 124

RESULT 8  
 ABC transport protein - *Synechocystis* sp. (strain PCC 6803)  
 N:Alternate names: protein sll0240  
 C:Species: *Synechocystis* sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 02-Feb-2001  
 C:Accession: S75100  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 S.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S75100  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <KAN>  
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAI1962.1; PID:g101  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homolo  
 C:Keywords: ATP; nucleotide binding; P-loop  
 F:17-210/Domain: ATP-binding cassette homology <ABC>  
 F:34-41/Region: nucleotide-binding motif A (P-loop)

Query Match 18.6%; Score 59; DB 2; Length 218;  
 Best Local Similarity 43.8%; Pred. No. 5.3;  
 Matches 14; Conservative 4; Mismatches 12; Indels 2; Gaps 2;

OY 13 RVNKGSTYFVDSVRCPCRRHPPHPPVFGF 44  
 :|||:::| | |||: | ||::| |  
 Db 52 RLDDGE-ITLGDROV-SCPQHYLPFHQIGIF 81

RESULT 9  
 homoserine dehydrogenase (hom) homolog - *Archaeoglobus fulgidus*  
 C:Species: *Archaeoglobus fulgidus*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 11-Jun-1999  
 C:Accession: G69366  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
 .; Goddek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 350, 364-370, 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Moese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: G69366  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-326 <KLE>  
 A:Cross-references: GB:AE001039; GB:AE000782; NID:g2689362; PIDN:AAB90303.1; PID:g264  
 C:Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology

Query Match 18.4%; Score 58.5; DB 2; Length 326;  
 Best Local Similarity 29.4%; Pred. No. 9.2;  
 Matches 15; Conservative 5; Mismatches 30; Indels 1; Gaps 1;

OY 8 IERVAVRGGEGTYFDVSVRNCPRHHFRRHPPVFGFCRADLFYDVALD 57  
 :|||:::| | |||: | ||::| |  
 Db 161 IESVKGIFNGTCNYILSRMEERLPYEHILKEAQELGVAEDPSYVEGID 211

RESULT 10  
 A54951  
 tissue factor pathway inhibitor-2 precursor - human  
 N:Alternate names: placental protein 5 (PP5)  
 C:Species: Homo sapiens (man)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: A54951; 155185; A34029; C34029; B34029  
 R:Spracher, C.A.; Kistler, W.; Mathews, S.; Foster, D.C.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 3353-3357, 1994  
 A:Title: Molecular cloning, expression, and partial characterization of a second huma  
 A:Reference number: A54951; MUID:94211862  
 A:Accession: A54951  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-225 <RES>  
 A:Cross-references: GB:L27624; NID:g441149; PIDN:AAA20094.1; PID:g441150  
 A:Experimental source: placenta

R.Miyagi, Y.: Koshikawa, N.; Yasumitsu, H.; Miyagi, E.; Hirahara, F.; Aoki, I.; Misugi, J. Biochem. 116, 939-942, 1994  
 A:Title: cDNA cloning and mRNA expression of a serine proteinase inhibitor secreted by *C. Superfamily: aspartate aminotransferase*  
 A:Reference number: 155185; MID:95204397  
 A:Accession: 155185  
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
 A:Molecule type: mRNA  
 A:Residues: 1-235 <RE2>  
 A:Cross-references: GB:D29992; NID:9484050; PIDN:BA06272.1; PID:9484051  
 A:Note: parts of this sequence, including the amino end of the mature protein, were detected  
 R:Buerzow, R.; Huhatala, M.L.; Bohn, H.; Virtanen, I.; Seppälä, M.  
 Biochem. Biophys. Res. Commun. 150, 483-490, 1988  
 A:Title: Purification and characterization of placental protein 5.  
 A:Reference number: A34029; MID:88106628  
 A:Accession: A34029  
 A:Molecule type: protein  
 A:Residues: 'A', 24-33, 'X', 35 <BU2>  
 A:Accession: C34029  
 A:Molecule type: protein  
 A:Residues: 47-50, 'X', 52-53 <BU2>  
 A:Accession: B34029  
 A:Molecule type: protein  
 A:Residues: 133, 'X', 135-137, 'X', 139-140, 'X', 142-144, 'X', 146 <BU3>  
 C:Genetics:  
 A:Gene: GDB:TFFP12  
 A:Cross-references: GDB:354485  
 C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor  
 C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-33/Domain: signal sequence #status predicted <SIG>  
 F:36-66/Domain: animal Kunitz-type proteinase inhibitor #status predicted <MNT>  
 F:66-149/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>  
 F:158-208/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>  
 F:36-66, 45-69, 61-82, 96-149, 106-130, 122-145, 158-208, 167-191, 183-204/Disulfide bonds: #str

Query Match 18.2%; Score 58; DB 2; Length 235;  
 Best Local Similarity 21.5%; Pred. No. 7.7;

Matches 14; Conservative 11; Mismatches 20; Indels 20; Gaps 2;

QY 1 ASFRVDRIEVARVR-----GGEGTYVDFSVRNCPR-----HHFPRHPN 40  
 DB 85 ACRRIEVPVCRILQVSDQCGSGSTPKYFNLSMTCGCFSGGCHRNRIENRFEAR 144

QY 41 VEGFC 45  
 DB 145 CMGFC 149

RESULT 11

AD0616  
 aspartate aminotransferase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh  
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
 A:Note: this species has also been called *Salmonella typhi*  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AD0616  
 R:Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AD0616  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-396 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05398.1; PID:g16502160; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: SRY1000  
 C:Superfamily: aspartate aminotransferase

Query Match 18.1%; Score 57.5; DB 2; Length 396;  
 Best Local Similarity 29.0%; Pred. No. 15;  
 Matches 18; Conservative 7; Mismatches 24; Indels 13; Gaps 3;

QY 6 DRIERVARVGGEGTYFV--DFSVRNCPR-----HHFPRHPNVGFC-----RADLFYD 52  
 DB 92 DKARRAQTPTGGTALRIADFLAKTPVKRWVSNPWNHKSVPNAGLEVREYAYTD 151

QY 53 VE 54  
 DB 152 AE 153

RESULT 12

T49147  
 nitrilase (EC 3.5.5.1) 1 [imported] - *Arabidopsis thaliana*  
 N:Alternate names: protein T10D17.100  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 03-Nov-2000  
 C:Accession: T49147; T52260; T52263  
 R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.  
 Submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z25017  
 A:Accession: T49147  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-346 <DAN>  
 A:Cross-references: EMBL:AL53865; GSPDB:GN00061; ATSP:T10D17.100  
 A:Experimental source: cultivar Columbia; BAC clone T10D17  
 R:Illstrand, H.; Bartling, D.; Weller, E.W.  
 Plant Mol. Biol. 36, 89-93, 1998  
 A:Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases,  
 A:Reference number: Z26007; MID:98145459  
 A:Accession: T52260  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-346 <HIL>  
 A:Cross-references: EMBL:Y07648; PIDN:CA06935.2  
 A:Experimental source: cultivar Columbia  
 R:Bartel, B.; Fink, G.R.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
 A:Title: Differential regulation of an auxin-producing nitrilase gene family in *Arabidopsis*  
 A:Reference number: Z24515; MID:8022831  
 A:Accession: T52263  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-346 <PAR>  
 A:Cross-references: EMBL:U38845; PIDN:AB05221.1  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Gene: ATSP:T10D17.100; ntl1, NIT1  
 A:Map position: 3  
 A:Introns: 44/1; 104/1; 202/1; 296/1  
 C:Superfamily: nitrilase  
 C:Keywords: hydrolase

Query Match 17.8%; Score 56.5; DB 2; Length 346;  
 Best Local Similarity 37.8%; Pred. No. 18;

Matches 14; Conservative 5; Mismatches 15; Indels 3; Gaps 1;

QY 18 EGTIVDFSVRNCPRHHFPRHPNVGFCRADLFYDVE 54  
 DB 227 EGGCFVLSACQFCQRKHFPDHP---YLFWDYDDKE 260

RESULT 13

S22398  
 nitrilase (EC 3.5.5.1) - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000  
 C:Accession: S22398  
 R:Bartling, D.; Seedorf, M.; Mithöfer, A.; Weller, E.W.



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:58 ; Search time 49.37 Seconds  
(without alignments)  
28.695 Million cell updates/sec

Title: US-09-730-379E-2

Perfect score: 318  
Sequence: 1 ASFRVDRIENRARRVGGEGT.....PNVGFQCADLFYVEALDL 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/6CTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	19.5	184	4	US-08-679-493A-213
2	58	18.2	213	2	US-08-796-850-2
3	58	18.2	235	1	US-08-147-710-2
4	58	18.2	235	1	US-08-458-090-2
5	58	18.2	235	2	US-08-457-887-2
6	58	18.2	235	3	US-08-817-145-3
7	55	17.3	446	4	US-09-457-046B-74
8	54.5	17.1	197	4	US-09-402-668-2
9	52	16.4	912	4	US-08-943-768-2
10	51.5	16.2	213	2	US-08-763-121-3
11	51.5	16.2	698	3	US-08-941-445A-11
12	51.5	16.2	1296	3	US-08-728-603-15
13	51	16.0	171	4	US-09-382-155-23
14	51	16.0	171	4	US-09-382-155-23
15	51	16.0	452	3	US-09-035-706-2
16	51	16.0	452	3	US-08-955-841-2
17	51	16.0	452	4	US-09-390-425-2
18	51	16.0	496	3	US-08-881-784-1
19	51	16.0	496	4	US-09-292-768-2
20	51	16.0	496	4	US-09-292-768-64
21	51	16.0	496	4	US-09-292-768-66
22	51	16.0	496	4	US-09-172-339-6
23	51	16.0	1433	2	US-08-365-486A-21
24	51	16.0	1433	4	US-09-123-708-4
25	51	16.0	1433	4	US-09-123-708-4
26	51	16.0	1433	4	US-08-880-342-21
27	51	16.0	1434	2	US-08-365-486A-19

28	51	16.0	1434	4	US-08-880-342-19	Sequence 19, Appl
29	51	16.0	1554	2	US-08-705-625-3	Sequence 3, Appl1
30	51	16.0	1554	4	US-09-010-998-6	Sequence 3, Appl1
31	51	16.0	1554	4	US-09-220-574-3	Sequence 6, Appl1
32	50.5	15.9	396	3	US-08-985-908-24	Sequence 24, Appl
33	50.5	15.9	1055	4	US-09-214-278-2	Sequence 2, Appl1
34	50.5	15.9	1212	4	US-09-214-278-3	Sequence 3, Appl1
35	50.5	15.9	1238	4	US-09-214-278-5	Sequence 5, Appl1
36	50.5	15.9	1257	3	US-08-611-729A-8	Sequence 8, Appl1
37	49.5	15.6	1065	2	US-08-400-159-8	Sequence 8, Appl1
38	49	15.4	278	2	US-09-260-283-2	Sequence 8, Appl1
39	48.5	15.3	219	2	US-08-379-057-30	Sequence 30, Appl
40	48.5	15.3	118	2	US-08-266-451B-24	Sequence 24, Appl
41	48.5	15.3	219	2	US-08-748-725-24	Sequence 5, Appl1
42	48.5	15.3	433	1	US-08-700-359-20	Sequence 20, Appl
43	48.5	15.3	453	4	US-09-362-831-5	Sequence 5, Appl1
44	48	15.1	316	1	US-08-403-634-4	Sequence 4, Appl1
45	48	15.1	316	1	US-08-403-634-31	Sequence 31, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-679-493A-213
Sequence 213, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: Taylor, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
PRIOR FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIOR FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIOR FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 213
LENGTH: 184
TYPE: PRT
ORGANISM: bloodfluke
US-08-679-493A-213

Query Match          19.5%; Score 62; DB 4; Length 184;
Best Local Similarity 30.4%; Pred. No: 0.88;
Matches 14; Conservative 7; Mismatches 17; Indels 8; Gaps 1;

QY 19 GTFVDFSVNCRPHHPRHNVFGCRA-----DLFYVEAL 56
      | | | | | | | | | | | | | | | | | | | | |
Db 91 GPHFNPNORHGRHGRHAGDLNIRVGSGVAKFDFVTIKGL 136

RESULT 2
US-08-796-850-2
Sequence 2, Application US/08796850
Patent No. 5981471
GENERAL INFORMATION:
APPLICANT: Papathanassiou, Adonia E
APPLICANT: Green, Shawn J.
TITLE OF INVENTION: Compositions and Methods for Inhibiting
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESS: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.A.
ZIP: 30303
COMPUTER READABLE FORM:
```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/796,850
7 FILING DATE:
8 CLASSIFICATION: 514
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Greene, Jamle L.
11 REGISTRATION NUMBER: 32,467
12 REFERENCE/DOCKET NUMBER: 05213-0290
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (404) 818-3700
15 TELEFAX: (404) 818-3799
16 INFORMATION FOR SEQ ID NO: 2:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 213 amino acids
19 TYPE: amino acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: protein
23 HYPOTHEICAL: NO
24 ANTI-SENSE: NO
25 FRAGMENT TYPE: N-terminal
26 ORIGINAL SOURCE:
27 ORGANISM: Homo sapiens
28
29 US-08-796-850-2

```

Query Match	18.2%	Score 58;	DB 2;	Length 213;
Best Local Similarity	21.5%	Pred. No. 3.5*		
Matches	14;	Conservative	11;	Mismatches 20; Indels 20; Gaps 2

  

QY	1	ASFRDRIERVARVR-----	GGGEATYEVDFSVRNCGR-----	HHPRRHN	40
		: : : : : : : : : : : :	: : : : : : : : : : : :		
Db	63	ACMRLEKPKVCRLQVSDVDCGEGSTERYTFNLSTCKEFTFGGCHRRRIENRFPDEAT			122
QY	41	VFGEFC	45		
Db	123	CMGFC	127		

RESULT 3  
 US-08-147-710-2  
 Sequence 2, Application US/08147710  
 Patent No. 5455338  
 GENERAL INFORMATION:  
 APPLICANT: Sprecher, Cindy A.  
 APPLICANT: Kistel, Walter  
 APPLICANT: Foster, Donald C.  
 TITLE OF INVENTION: NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND  
 TITLE OF INVENTION: METHODS RELATING THERETO  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Zymogenetics, Inc.  
 STREET: 4225 Roosevelt Way, N.E.  
 CITY: Seattle  
 STATE: WA  
 COUNTRY: USA  
 ZIP: 98105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/147,710  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, Galy E  
 REGISTRATION NUMBER: 31-684

```

: REFERENCE/DOCKET NUMBER: 93-14
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-547-8080 ext 322
: TELEFAX: 206-548-2329
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 235 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
US-08-147-710-2

```

```

OY      1 ASFRVDRIERVARVR-----GGEGRYFVDFVSFRNCDP-----HHFPRHNP 40
Db      85 ACWRREKPKVCRGLQVSDVDDGEGSTETKFFFLSLSMTCCKFFSGGCHRRRIENRPFDEAT 144
OY      41 VFGFC 45
Db      145 CMGFC 149

```

```

RESULT      4
US-08-458-090-2
; Sequence 2, Application US/08458090
; Patent No. 5728674
; GENERAL INFORMATION:
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Kistel, Walter
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
; TITLE OF INVENTION: METHODS RELATING THERETO
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.090
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-14D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-090-2

```

Query Match	18.2%;	Score 58;	DB 1;	Length 235;
Best Local Similarity	21.5%;	Pred. No. 3.9;		
Matches	14;	Conservative	11;	Mismatches
			20;	Indels
				20;
				Gaps
				2;



Query Match	17.3%	Score 55	DB 4	Length 446
Best Local Similarly	47.8%	Pred. No. 20		
Matches 11; Conservative	2;	Mismatches 10;	Indels 0;	Gaps 0

QY 18 EGYTFVDFSVRNCPRHHFPRHPN 40  
 11 111 111 1 : 1 :  
 Db 92 EGAIFVDARVNNCPLETFELKCPD 114

```

RESULT
8
US-09-402-668-2
: Sequence 2, Application US/09402668
: Patent No. 6172030
: GENERAL INFORMATION:
: APPLICANT: WADA, Yasunao
: APPLICANT: KASAI, Miyuki
: APPLICANT: SHIKATA, Shitsuo
: APPLICANT: SUZUMATSU, Aetsushi
: APPLICANT: KOIKE, Kenzo
: APPLICANT: HATADA, Yoji
: APPLICANT: KOBAYASHI, Tohru
: APPLICANT: ITO, Susumu
: APPLICANT: TSUMADORI, Masaki
: TITLE OF INVENTION: Detergent Composition
: FILE REFERENCE: 2173-0116P
: CURRENT APPLICATION NUMBER: US/09/402,668
: PRIORITY FILING DATE: 1998-10-08
: PRIOR APPLICATION NUMBER: 9-091142 JAPAN
: PRIOR FILING DATE: 1997-04-09
: PRIOR APPLICATION NUMBER: 9-242736 JAPAN
: PRIOR FILING DATE: 1997-09-08
: PRIOR APPLICATION NUMBER: PCT/US98/01613
: PRIOR FILING DATE: 1998-04-09
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 197
: TYPE: PRF
: ORGANISM: Bacillus sp.
: OTHER INFORMATION: Strain: KSM-P15
: US-09-402-668-2

```

```

QY      2  SFRVDRIERVARVRCGGCTGYFYDESVRNC 30
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      121 NFRADDICKLVNRQNGT-TYKVVNNVNE 148
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1
Query Match 17.1%; Score 54.5; DB 4; length 197;
Best Local Similarity 44.8%; Pred. 9.2;

```

```

RESULT          9
US-08-943-768-2
: Sequence 2, Application US/08943768C
: Patent No. 6238881
: GENERAL INFORMATION:
: APPLICANT: Hart, Matthew J.
: TITLE OF INVENTION: No. 6238881el Nucleic Acids and Polypeptides Related to a
: TITLE OF INVENTION: Guanine Exchange Factor of RHO GTPase
: FILE REFERENCE: 1023-US
: CURRENT APPLICATION NUMBER: US/08/943,768C
: CURRENT FILING DATE: 1997-10-06
: EARLIER APPLICATION NUMBER: 60/029,979
: EARLIER FILING DATE: 1996-11-06
: NUMBER OF SEQ. ID NOS: 12
: SOFTWARE: Patentln Ver. 2.0
: SEQ. ID NO 2
: LENGTH: 912
: TYPE: PRT
: ORGANISM: Human pl15 GEF-Rho
: US-08-943-768-2

```

Query Match	16.48;	Score 52;	DB 4;	Length 912;
Best Local Similarity	44.08;	Pred. No. 1.1e+02;		
Matches 11; Conservative	3;	Mismatches 11;	Indels 0;	Gaps 0;

```
QY 11 VARVRGEGTYFVDFSVRNCPRHHF 35
      :|| | |::| | | | |
Db 497 LARFDGAEGSWFQKISSRRCRSQSF 521
```

```

RESULT 10
US-08-763-121-3
: Sequence 3, Application US/08/763121
: Patent No. 5916764
:
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Goll, Surya K.
: TITLE OF INVENTION: HUMAN PHOSPHATIDYLINOSITOL SYNTHASE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: US
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/763,121
: FILING DATE: Filed Herewith
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0165 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 213 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 1620879
:
US-08-763-121-3

```

```

QY      14 VRGEGCITYFVDSVANCPRHHPRRHPNVFGFCRA-DLEY 51
          : : : : : : : : : : : : : : : : : : : :
ob      114 VRGESHKMIIDLSCGNPVLRIYTSRPALEFTLCAGNELEY 152
          : : : : : : : : : : : : : : : : : : : :
          Matches 13; Conservative 5; Mismatches 20; Indels 1; Gaps 1

```

US-RESULT 11  
 08-98-941-445A-11  
 ; Sequence 11, Application US/08941445A  
 ; Patent No. 6107060  
 ;  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keeling, Peter  
 ; APPLICANT: Guan, Heping  
 ; TITLE OF INVENTION: Starch Encapsulation  
 ;  
 ; NUMBER OF SEQUENCES: 37  
 ;  
 ; CORRESPONDENCE ADDRESS: .

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,445A  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,855  
FILING DATE: 30-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Winner, Ellen P  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 89-97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 698 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-941-445A-11

Query Match 16.2%; Score 51.5; DB 3; Length 698;  
Best Local Similarity 36.5%; Pred. No. 95;  
Matches 19; Conservative 5; Mismatches 15; Indels 13; Gaps 4;

QY 8 IEVAVRG--GEGTYF-----VDFSVRNC--RHHPRHVFGCRADL 49  
DB 261 VRRRYVAGDSVYTFHSTYIDGVDFVEVAPPRH--RHNNYGGRLDI 309

RESULT 12  
US-08-728-603-15  
Sequence 15, Application US/08728603  
Patent No. 6093806  
GENERAL INFORMATION:  
APPLICANT: Caesarman, Ethel  
TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED  
TITLE OF INVENTION: HERPESVIRUS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: USA  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,603  
FILING DATE: 10-OCT-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BRAMAN, SUSAN J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 19603/720

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-263-1636  
TELEFAX: 716-263-1600  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-728-603-15

Query Match 16.2%; Score 51.5; DB 3; Length 1296;  
Best Local Similarity 29.4%; Pred. No. 1,9e+02;  
Matches 15; Conservative 8; Mismatches 23; Indels 5; Gaps 2;

QY 3 FRVDRIEVARVRG--EGTYFVDFSVRNCPRHHPRP--NVFGCRAD 48  
DB 1193 YEQDALEYLRQGEITLTLYHGNADETLPRARYPRPPTGNSVAGLTSSD 1243

RESULT 13  
US-09-382-155-23  
Sequence 23, Application US/09382155B  
Patent No. 6160095  
GENERAL INFORMATION:  
APPLICANT: CHAUDHARY, PREET M  
TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND APOPTOSIS  
TITLE OF INVENTION: PATHWAYS AND METHODS OF USING THE SAME  
FILE REFERENCE: Chaudhary  
CURRENT APPLICATION NUMBER: US/09/382,155B  
EARLIER FILING DATE: 1999-08-24  
EARLIER APPLICATION NUMBER: 09/074,044  
EARLIER FILING DATE: 1998-05-07  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 23  
LENGTH: 171  
TYPE: PRT  
ORGANISM: Equine Herpesvirus  
US-09-382-155-23

Query Match 16.0%; Score 51; DB 4; Length 171;  
Best Local Similarity 35.3%; Pred. No. 22;  
Matches 12; Conservative 8; Mismatches 8; Indels 6; Gaps 2;

QY 3 FRVDRIEVARVRGEGTYFVDFSVRNCPRHHP 36  
DB 61 FRVGRDLIRIRFG--QTWIPD-----SCPRTYMP 88

RESULT 14  
US-09-074-044A-23  
Sequence 23, Application US/09074044A  
Patent No. 6207458  
GENERAL INFORMATION:  
APPLICANT: CHAUDHARY, PREET M  
TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND  
TITLE OF INVENTION: APOPTOSIS PATHWAYS AND METHODS OF USING THE SAME  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS  
STREET: 2405 GRAND BLVD., SUITE 400  
CITY: KANSAS CITY  
STATE: MISSOURI  
COUNTRY: USA  
ZIP: 64108  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Search completed: May 24, 2002, 16:48:59  
Job time: 284 sec

; INFORMATION FOR SEQ ID NO: 2:

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:02 ; Search time 136.41 Seconds  
(without alignments)  
47.227 Million cell updates/sec

Title: US-09-730-379E-2  
Perfect score: 318  
Sequence: 1 ASFRVRIERHVARVRGEGT.....PNVFGFCRADLFYVEALDL 58

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_032802:\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	168.5	53.0	155	22	ABG18910
2	64	20.1	573	21	AAAG29545
3	64	20.1	576	21	AAAG29544
4	64	20.1	604	21	AAAG29543
5	60	18.9	203	21	AAAB43068
6	60	18.9	203	22	AAAM39350
7	59.5	18.7	14	22	ABBS2112
8	58	18.2	188	22	AAU02970
9	58	18.2	210	22	AAU02970
10	58	18.2	211	22	AAU02969
11	58	18.2	213	19	AAW61536

12	58	18.2	235	16	AAAR74977
13	58	18.2	235	22	AAAB76856
14	57.5	18.1	179	22	AAAM80018
15	57	17.9	315	22	ABG01233
16	56.5	17.8	193	21	AAAG23444
17	56.5	17.8	193	21	AAAG23443
18	56.5	17.8	224	21	AAAG23442
19	56	17.6	430	22	ABBB6850
20	55.5	17.5	296	22	ABBB60055
21	55	17.3	104	21	AAAG61693
22	55	17.3	179	22	ABBI1529
23	55	17.3	179	22	AAAM41136
24	55	17.3	345	21	AAV74701
25	55	17.3	709	10	AAAP91934
26	54.5	17.1	197	19	AAAM83014
27	54.5	17.1	197	19	AAW77412
28	54.5	17.1	246	20	AAW99400
29	54.5	17.1	429	22	AAU00392
30	54	17.0	452	22	AAAB84202
31	53.5	16.8	123	21	AAAG40911
32	53.5	16.8	193	21	AAAG40910
33	53.5	16.8	218	21	AAAG40909
34	53.5	16.8	351	19	AAW72022
35	53.5	16.8	466	19	AAW72230
36	53.5	16.8	523	19	AAW72229
37	53.5	16.8	610	19	AAW72228
38	53.5	16.8	649	19	AAW72097
39	53.5	16.8	2175	22	ABBS59839
40	53	16.7	80	22	AAAM88730
41	53	16.7	144	22	ABG06951
42	52.5	16.5	445	22	ABG27667
43	52.5	16.5	1719	22	ABBB70837
44	52	16.4	134	22	ABBB70115
45	52	16.4	404	22	ABG04912

## ALIGNMENTS

RESULT 1  
ABG18910  
ID ABG18910 standard; Protein: 155 AA.  
AC ABG18910;  
DT 18-FEB-2002 (first entry)  
DE Novel human diagnostic protein #18901.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Yang YP;  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB; AAS83097.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess



PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148664.  
PR 16-AUG-1999; 99US-0149366.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155658.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156599.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 20.1%; Score 64; DB 21; Length 573;  
Best Local Similarity 41.3%; Pred. No. 4;  
Matches 19; Conservative 3; Mismatches 12; Indels 12; Gaps 3;

Oy 6 DRIERARVGGEGTYFVDSVNCPRHRHPRHPNVEGCRADLEY 51  
Db 464 dgIeneeqmkakeglfYpfs-----hfr--pnkl---rxdcfy 497

RESULT 3  
AAC29544  
ID AAC29544 standard; Protein; 576 AA.  
XX AAC29544;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35171.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132853.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134418.  
PR 14-MAY-1999; 99US-01344219.  
PR 14-MAY-1999; 99US-0134421.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.



PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 20.1%; Score 64; DB 21; Length 576;  
Best Local Similarity 41.3%; Pred. No. 4.1;  
Matches 19; Conservative 3; Mismatches 12; Indels 12; Gaps 3;

OY 6 DRIEVARVGGEGTYFVDSVRNCPRHHFPRHPVFCGRADLPY 51  
Db 467 dgieneqmkakegcltlypfis-----hfp--pnkl---tkdcify 500

RESULT 4  
ID AAG29543  
AC AAG29543; standard; Protein; 604 AA.  
AC AAG29543;  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35170.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145819.  
PR 28-JUL-1999; 99US-0145591.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 20.1%; Score 64; DB 21; Length 604;  
Best Local Similarity 41.3%; Pred. No. 4.3;  
Matches 19; Conservative 3; Mismatches 12; Indels 12; Gaps 3;  
  
OY 6 DRIEVARVAGGEGTYFVDSVRCNCPRHPPRHPPNFGFCRADLEY 51  
Db 495 dyleneegmkakegllfyps-----htp--pnkl---rkdcfy 528  
  
RESULT 5  
AAB43068  
ID AAB43068 standard; Protein; 203 AA.  
XX  
AC AAB43068;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2832 polypeptide sequence SEQ ID NO:5664.  
XX  
KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000MO-US08621.  
XX  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
PI WPI: 2000-602362/57.  
XX  
DR N-PSDB; AAC77277.  
XX  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
XX PT useful for treating e.g. cancers, proliferative disorders,  
XX PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 11; Page 4839-4840; 5507pp; English.  
XX  
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
CC CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;



PS Example: Page 28; 162pp; English.

XX The invention relates to methods for the screening, diagnosis and

CC prognosis of Alzheimer's disease. The methods involve the detection

CC of Alzheimer's Disease-Associated Features (AFs) and Alzheimer's

CC Disease-Associated Protein Isoforms (APIs) in cerebrospinal fluid,

CC serum or plasma. The abundance of the AFs and APIs is then

CC normalised to an Expression Reference Protein Isoform (ERPI) in

CC order to determine whether a patient is suffering from, or has

CC a predisposition to, Alzheimer's disease. The relative abundance of

CC the AFs and APIs correlates with the severity of Alzheimer's Disease.

CC The present sequence is a peptide produced from an API by proteolysis.

XX

XX Sequence 14 AA;

SQ

Query Match 18.7%; Score 59.5; DB 22; Length 14;

Best Local Similarity 92.9%; Pred. No. 0.22;

Matches 13; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 16 GGEGT-YFVDFSVR 28

Db 1 ggegtgfyvdfsvr 14

|||||

RESULT 8

AAU02979

ID AAU02979 standard; Protein; 188 AA.

XX

XX AAU02979;

AC

XX

DT 12-SEP-2001 (first entry)

XX

DE Angiotensin converting enzyme (ACEV) splice variant protein #79.

XX

KM Angiotensin converting enzyme splice variant; ACEV; interleukin 6;

KM granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

KM platelet-derived endothelial cell growth factor; cardiovascular disease;

KM cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;

KM vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;

KM myocardial infarction; coronary arterial thrombosis; renal disease;

KM diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;

KM multiple sclerosis; immune complex nephritis; deep vein thrombosis;

KM noncardioidic pulmonary granulomatous disease; endothelial abnormality;

KM vascular disorder; asbestosis.

XX

OS Homo sapiens.

XX

PN WO200136632-A2.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000MO-IL00766.

XX

PR 17-NOV-1999; 99IL-0132978.

PR 10-DEC-1999; 99IL-0133455.

XX

PA (COMP-) COMPUGEN LTD.

XX

PI Levine Z, David A, Azar I, Khosravi R, Bernstein J;

XX

DR WPI: 2001-336004/35.

DR N-PSDB; AAS06079.

XX

PT Novel alternative splicing variants e.g. variant of angiotensin

PT converting enzyme (ACEV), useful in identifying candidate compounds

PT capable of binding to the variant and to detect anti-variant antibodies

PT

XX

PS Claim 4; Fig 79; 519pp; English.

XX

CC The sequence represents an angiotensin converting enzyme splice variant

CC (ACEV) polypeptide. The polypeptides of the invention include variants of

CC granulocyte colony stimulating factor receptor, glucagon, interleukin 6,

CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase

CC inhibitor 1C, cellular tumour antigen P53, and vasoactive intestinal

CC polypeptide receptor 2. The polypeptides and their associated nucleic

CC acids are useful for identification of variant sequences and detection of

CC candidate compounds capable of binding the molecules. The sequences of

CC the invention can be used in the treatment and diagnosis of various

CC disorders including cardiovascular diseases such as arteriosclerosis,

CC myocardial infarction and coronary arterial thrombosis, renal diseases

CC such as diabetic nephropathy, muscular diseases such as hypertrophy,

CC immune disorders such as immune complex nephritis, multiple sclerosis,

CC cancer, sarcoidosis, noncardioidic pulmonary granulomatous diseases such

CC as asbestosis and vascular pathologies involving an endothelial

CC abnormality such as deep vein thrombosis.

XX

XX Sequence 188 AA;

SQ

Query Match 18.2%; Score 58; DB 22; Length 188;

Best Local Similarity 21.5%; Pred. No. 7.2;

Matches 14; Conservative 11; Mismatches 20; Indels 20; Gaps 2;

QY 1 ASFRVDRIERVAVR-----GGEGTYFVDFSVRNCPR-----HHFRHPN 40

Db 85 acwrlkxvprctqysvddqcegstekeyifnlsmtcckfsgyghnrnriemfpdeat 144

|||||

QY 41 VFGFC 45

Db 145 cmgfc 149

|||||

RESULT 9

AAU02970

ID AAU02970 standard; Protein; 210 AA.

XX

XX AAU02970;

AC

XX

DT 12-SEP-2001 (first entry)

XX

DE Angiotensin converting enzyme (ACEV) splice variant protein #70.

XX

KM Angiotensin converting enzyme splice variant; ACEV; interleukin 6;

KM granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

KM platelet-derived endothelial cell growth factor; cardiovascular disease;

KM cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;

KM vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;

KM myocardial infarction; coronary arterial thrombosis; renal disease;

KM diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;

KM multiple sclerosis; immune complex nephritis; deep vein thrombosis;

KM noncardioidic pulmonary granulomatous disease; endothelial abnormality;

KM vascular disorder; asbestosis.

XX

OS Homo sapiens.

XX

PN WO200136632-A2.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000MO-IL00766.

XX

PR 17-NOV-1999; 99IL-0132978.

PR 10-DEC-1999; 99IL-0133455.

XX

PA (COMP-) COMPUGEN LTD.

XX

PI Levine Z, David A, Azar I, Khosravi R, Bernstein J;

XX

DR WPI: 2001-336004/35.

DR N-PSDB; AAS06070.

XX

PT Novel alternative splicing variants e.g. variant of angiotensin

PT converting enzyme (ACEV), useful in identifying candidate compounds

PT capable of binding to the variant and to detect anti-variant antibodies

PT

PT  
XX  
PS  
XX  
XX  
Claim 4; Fig 70; 519pp; English.

CC The sequence represents an angiotensin converting enzyme splice variant  
CC (ACEV) polypeptide. The polypeptides of the invention include variants of  
CC granulocyte colony stimulating factor receptor, glucagon, interleukin 6,  
CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase  
CC inhibitor 1C, cellular tumour antigen p53, and vasodilator intestinal  
CC polypeptide receptor 2. The polypeptides and their associated nucleic  
CC acids are useful for identification of variant sequences and detection of  
CC candidate compounds capable of binding the molecules. The sequences of  
CC the invention can be used in the treatment and diagnosis of various  
CC disorders including cardiovascular diseases such as arteriosclerosis,  
CC myocardial infarction and coronary arterial thrombosis, renal diseases  
CC such as diabetic nephropathy, muscular diseases such as hypertrophy,  
CC immune disorders such as immune complex nephritis, multiple sclerosis,  
CC cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such  
CC as asbestosis and vascular pathologies involving an endothelial  
CC abnormality such as deep vein thrombosis.

XX  
SQ Sequence 210 AA;

Query Match 18.2%; Score 58; DB 22; Length 210;  
Best Local Similarity 21.5%; Pred. No. 8.2;  
Matches 14; Conservative 11; Mismatches 20; Indels 20; Gaps 2;

OY 1 ASFRVDRIERVARVR-----GGEGTYFVDFSVRNCPR-----HHRPRHPN 40

DB 85 acwrliekvkvcrlysvdqcgstekyffnlsmctckffsgghrnrlnrpfdeat 144

OY 41 VFGFC 45

DB 145 cmgfc 149

RESULT 10

AAU02969  
ID AAU02969 standard; protein; 211 AA.

AC AAU02969;

DT 12-SEP-2001 (first entry)

DE Angiotensin converting enzyme (ACEV) splice variant protein #69.

XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;  
KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;  
KW platelet-derived endothelial cell growth factor; cardiovascular disease;  
KW cellular tumour antigen p53; cyclin-dependent kinase inhibitor 1C;  
KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;  
KW myocardial infarction; coronary arterial thrombosis; renal disease;  
KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;  
KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;  
KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;  
KW vascular disorder; asbestosis.

XX Homo sapiens.

OS  
XX  
XX WO200136632-A2.

PN  
XX  
XX 25-MAY-2001.

PD  
XX  
XX 17-NOV-2000; 2000WO-IL00766.

PF  
XX  
XX 17-NOV-1999; 99TL-0132978.

PR  
XX  
XX 10-DEC-1999; 99TL-0133455.

XX (COMP-) COMPUGEN LTD.

PA  
XX Levine Z, David A, Azar I, Khostravl R, Bernstein J;  
PI  
XX

DR WPI; 2001-336004/35.  
DR N-PSDB; AAS06069.

PT Novel alternative splicing variants e.g. variant of angiotensin

PT converting enzyme (ACEV), useful in identifying candidate compounds

PT capable of binding to the variant and to detect anti-variant antibodies

PS Claim 4; Fig 69; 519pp; English.

XX The sequence represents an angiotensin converting enzyme splice variant  
XX (ACEV) polypeptide. The polypeptides of the invention include variants of  
XX granulocyte colony stimulating factor receptor, glucagon, interleukin 6,  
XX platelet-derived endothelial cell growth factor, cyclin-dependent kinase  
XX inhibitor 1C, cellular tumour antigen p53, and vasodilator intestinal  
XX polypeptide receptor 2. The polypeptides and their associated nucleic  
XX acids are useful for identification of variant sequences and detection of  
XX candidate compounds capable of binding the molecules. The sequences of  
XX the invention can be used in the treatment and diagnosis of various  
XX disorders including cardiovascular diseases such as arteriosclerosis,  
XX myocardial infarction and coronary arterial thrombosis, renal diseases  
XX such as diabetic nephropathy, muscular diseases such as hypertrophy,  
XX immune disorders such as immune complex nephritis, multiple sclerosis,  
XX cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such  
XX as asbestosis and vascular pathologies involving an endothelial  
XX abnormality such as deep vein thrombosis.

XX  
SQ Sequence 211 AA;

Query Match 18.2%; Score 58; DB 22; Length 211;  
Best Local Similarity 21.5%; Pred. No. 8.3;  
Matches 14; Conservative 11; Mismatches 20; Indels 20; Gaps 2;

OY 1 ASFRVDRIERVARVR-----GGEGTYFVDFSVRNCPR-----HHRPRHPN 40

DB 85 acwrliekvkvcrlysvdqcgstekyffnlsmctckffsgghrnrlnrpfdeat 144

OY 41 VFGFC 45

DB 145 cmgfc 149

RESULT 11

AAW61536  
ID AAW61536 standard; protein; 213 AA.

AC AAW61536;

DT 06-NOV-1998 (first entry)

DE Human tissue factor pathway inhibitor-2 (TFPI-2).

XX Human tissue factor pathway inhibitor; TFPI; TFPI-2; cell proliferation;  
KW angiogenesis-related disease; cancer; arthritis; macular degeneration;  
KW diabetic retinopathy; placental protein 5.

XX Homo sapiens.

OS  
XX  
XX Key Location/Qualifiers

FT Disulfide-bond 14..64

FT Disulfide-bond 23..47

FT Disulfide-bond 39..60

FT Disulfide-bond 74..127

FT Disulfide-bond 84..108

FT Disulfide-bond 100..123

FT Disulfide-bond 136..186

FT Disulfide-bond 145..169

FT Disulfide-bond 161..182

XX WO9834634-A1.

PN  
XX  
XX 13-AUG-1998.



```
XX Example 1; Page 260-261; 436pp; English.
PS
XX
CC The present invention describes immunogenic portions of lung tumour-
CC associated proteins (I) and the nucleic acids (NAs) that encode them.
CC (I) have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patients own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (I) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF68878 and
CC AAB76848 to AAB76878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 235 AA;

Query Match 18.2%; Score 58; DB 22; Length 235;
Best Local Similarity 21.5%; Pred. No. 9.4;
Matches 14; Conservative 11; Mismatches 20; Indels 20; Gaps 2;

QY 1 ASRRVDRIERVARVR-----GGEGTYFVDFSVRNCPR-----HHRPRHPN 40
   | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 85 acvriekvpkvcrlgysvddqcgstekeyffnissmtcekfifsgchrrnlenrfdeat 144

QY 41 VVEGEC 45
   |||
Db 145 cmgfc 149

RESULT 14
AAB80018
ID AAB80018 standard; Protein; 179 AA.
XX
AC AAB80018;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3664.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
```

```
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB: AAK53151.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PR useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 409; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activating/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 179 AA;

Query Match 18.1%; Score 57.5; DB 22; Length 179;
Best Local Similarity 31.1%; Pred. No. 8;
Matches 14; Conservative 4; Mismatches 12; Indels 15; Gaps 2;

QY 8 IERVAVRGEGTYFVDFSVRNCRRHHRPRVVFGRADLFYD 52
   : | | | | | | | | | | | | | | | : : :
Db 33 lstrgtvpgyrga-----giprhpsr-----cclqylfn 62

RESULT 15
ABG01233
ID ABG01233 standard; Protein; 315 AA.
XX
AC ABG01233;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #1224.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS65420.
```

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX  
 PS Claim 20; SEQ ID No 31592; 103pp; English.

XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. Abg00010-ABg3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SO Sequence 315 AA;

Query Match 17.9%; Score 57; DB 22; Length 315;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 13 RVRGGEGTFVDFSV 27  
 :|||||||::||  
 Db 247 kvrgsegtylhlsv 261

Search completed: May 24, 2002, 16:48:03  
 Job time: 243 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:02:49 ; Search time 107.9 Seconds  
(without alignments)  
14.430 Million cell updates/sec

Title: US-09-730-379E-3  
Perfect score: 58  
Sequence: 1 GPRPFHCRO 9

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	77.6	216	13	Q918W0
2	45	77.6	431	10	Q91ZU2
3	45	77.6	481	13	Q90X58
4	45	77.6	483	4	Q9H2S9
5	45	77.6	513	13	Q918V9
6	45	77.6	522	13	Q918W2
7	45	77.6	533	11	Q9Z222
8	45	77.6	545	4	Q96UP3
9	45	77.6	563	13	Q90W82
10	45	77.6	725	4	Q96S24
11	44	75.9	428	13	Q90WJ6
12	42	72.4	441	10	Q02056
13	42	72.4	553	4	Q9UWM2
14	42	72.4	654	4	Q96CP9
15	42	72.4	707	10	Q40054
16	42	72.4	784	4	O00146

17	41	70.7	588	13	Q9DG49	Q9dg49 brachydanio
18	41	70.7	599	13	Q9PSH2	Q9psH2 gallus gall
19	40	69.0	158	5	Q9U509	Q9u509 manduca sex
20	40	69.0	227	10	Q43479	Q43479 hordeum vul
21	40	69.0	411	13	Q9W747	Q9w747 brachydanio
22	40	69.0	498	4	Q9H8L4	Q9h8L4 homo sapien
23	40	69.0	619	4	Q96K58	Q96k58 homo sapien
24	40	69.0	720	6	Q9N003	Q9n003 macaca fasc
25	39	67.2	67	11	Q64493	Q64493 mus musculu
26	39	67.2	101	2	Q49692	Q49692 mycobacteri
27	39	67.2	274	11	Q91VL6	Q91vL6 mus musculu
28	39	67.2	400	5	Q9W1E6	Q9w1E6 drosophila
29	39	67.2	422	11	Q9ET64	Q9et64 ratius norv
30	39	67.2	446	4	Q96W79	Q96w79 homo sapien
31	39	67.2	742	4	Q9H0M5	Q9h0M5 homo sapien
32	39	67.2	822	11	Q88961	Q88961 ratius norv
33	39	67.2	1043	11	Q9WTW1	Q9wtW1 ratius norv
34	38	65.5	179	13	Q42379	Q42379 brachydanio
35	38	65.5	465	2	Q910Y1	Q910Y1 streptomyce
36	38	65.5	896	4	Q9C0E4	Q9c0E4 homo sapien
37	38	65.5	1891	5	Q77275	Q77275 drosophila
38	38	65.5	1893	5	Q9W4J1	Q9w4J1 drosophila
39	38	65.5	1920	5	Q46205	Q46205 drosophila
40	37	63.8	98	6	Q9WYT1	Q9wyt1 oryctolagus
41	37	63.8	138	13	Q13100	Q13100 xenopus lae
42	37	63.8	370	11	Q921X0	Q921X0 mus musculu
43	37	63.8	392	13	Q90XA1	Q90xa1 brachydanio
44	37	63.8	402	4	Q96CLO	Q96cLO homo sapien
45	37	63.8	408	4	Q96H86	Q96h86 homo sapien

#### ALIGNMENTS

RESULT 1  
ID Q918W0 PRELIMINARY; PRT; 216 AA.  
AC Q918W0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE EOS (FRAGMENT).  
OS Raja eglanteria.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squalae; Hymnosqualeae; Pristionotae; Batoidae;  
OC Rajiformes; Rajidae; Raja.  
OX NCBI\_TaxID=35514;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20318730; PubMed=10861066;  
RA Haire R.N., Miracle A.L., Rast J.P., Litman G.W.;  
RT "Members of the Ikaros gene family are present in early representative  
vertebrates.";  
RL J. Immunol. 165:306-312(2000).  
DR EMBL: AF163849; AAF87272.1; -;  
DR HSSP: P15822; 1BBO.  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00096; Zf-C2H2; 4.  
DR SMART: SM00355; Znf-C2H2; 4.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 4.  
KW DNA-binding; Metal-binding; zinc-finger.  
FT NON\_TER 1  
FT SEQUENCE 216 216  
SQ SEQUENCE 216 AA; 4D4BF896ECD74E67 CRC64;

Query Match 77.6%; Score 45; DB 13; Length 216;

Best local Similarity 77.8%; Pred. No. 1.4;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
| | | | | | |

Db 114 GERPFHCN0 122

RESULT 2

ID 09LZU2

PRELIMINARY;

PRT; 431 AA.

AC 09LZU2; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE PHOSPHOLIPASE-LIKE PROTEIN.

GN F1612.90.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_Taxid=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Jordan N., Bangert S., Medelmann R., Voss H., Unseid M., Mewes H.W.,

RA Rudd S., Lemcke K., Meyer K.F.X., Queller F., Salanoubat M.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162459; CAB82812.1; -

SQ SEQUENCE 431 AA; 48980 MW; 567216AE7FED3D7 CRC64;

Query Match

Best Local Similarity 77.6%; Score 45; DB 10; Length 431;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GERPFHC 7

Db 259 GPRPFHC 265

RESULT 3

ID 090X58

PRELIMINARY;

PRT; 481 AA.

AC 090X58;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE TRANSCRIPTION FACTOR HELIOS.

OS Ambystoma mexicanum (Axolotl).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;

OC Ambystoma.

OX NCBI\_Taxid=8296;

RN [1]

RP SEQUENCE FROM N.A.

RA Durand C., Kerfourn F., Charlemagne J., Fellah J.S.;

RT "Structure and expression of Helios in the Mexican axolotl.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF431972; AAL27462.1; -

SQ SEQUENCE 481 AA; 53190 MW; 730488CF0B22202 CRC64;

Query Match

Best Local Similarity 77.6%; Score 45; DB 13; Length 481;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHC 9

Db 136 GERPFHCN0 144

RESULT 4

ID 09H2S9

PRELIMINARY;

PRT; 483 AA.

AC 09H2S9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE ZINC FINGER TRANSCRIPTION FACTOR EOS.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20556271; PubMed=10978333;

RA Perdomo J., Holmes M., Chong B., Crossley M.;

RT "Eos and Pegasus, Two Members of the Ikaros Family of Proteins with

RT Distinct DNA Binding Activities.";

RL J. Biol. Chem. 275:38347-38354(2000).

DR EMBL; AF230809; AAG39221.1; -

DR HSSP; P15822; 1BBO.

DR InterPro; IPR000822; Znf-C2H2.

DR Pfam; PF00096; zf-C2H2; 6.

DR SMART; SM00355; Znf\_C2H2; 6.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.

DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.

KW DNA-binding; Zinc-finger.

SQ SEQUENCE 483 AA; 53219 MW; C30B0270709C34F9 CRC64;

Query Match

Best Local Similarity 77.6%; Score 45; DB 4; Length 483;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GERPFHC 9

Db 81 GERPFHCN0 89

RESULT 5

ID 091BV9

PRELIMINARY;

PRT; 513 AA.

AC 091BV9;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE AIOLOS.

OS Raja eglanteria.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Squalae; Hypnosqualae; Pristiogadae; Batoidae;

OC Rajiformes; Rajidae; Raja.

OX NCBI\_Taxid=33514;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20318730; PubMed=10861066;

RA Haire R.N., Miracle A.L., Rast J.P., Litman G.W.;

RT "Members of the Ikaros gene family are present in early representative

RT vertebrates.";

RL J. Immunol. 165:306-312(2000).

DR EMBL; AF163850; AAF87273.1; -

DR HSSP; P15822; 1BBO.

DR InterPro; IPR000822; Znf-C2H2.

DR Pfam; PF00096; zf-C2H2; 5.

DR SMART; SM00355; Znf\_C2H2; 6.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.

DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.

KW DNA-binding; Metal-binding; Zinc-finger.

SQ SEQUENCE 513 AA; 56940 MW; 9D1F5CEB25B6D110 CRC64;

Query Match

Best Local Similarity 77.6%; Score 45; DB 13; Length 513;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHC 9

Db 1111111 1

```

Db      145 GERPFHCNO 153

RESULT 6
ID      0918W2      PRELIMINARY:      PRT:      522 AA.
AC      0918W2;
DT      01-OCT-2000 (TREMBlrel. 15, Created)
DT      01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      HELIOS.
OS      Raja eglanteria.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC      Elasmobranchii; Squalae; Hyposqualeae; Pristiorajae; Batoidae;
OC      Rajiformes; Rajidae; Raja.
OX      NCBI_TaxID=33514;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20318730; PubMed=10861066;
RA      Haire R.N., Miracle A.L., Rast J.P., Litman G.W.;
RT      "Members of the Ikaros gene family are present in early representative
RT      vertebrates." 165:306-312(2000).
RL      J. Immunol. 165:306-312(2000).
DR      EMBL; AF163847; AAF87270.1; -.
DR      HSSP; P15822; 1BBO.
DR      InterPro; IPR000822; znf-C2H2.
DR      Pfam; PF00096; zf-C2H2; 5.
DR      SMART; SM00355; Znf_C2H2; 6.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR      PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW      DNA-binding; Metal-binding; Zinc-finger.
SQ      SEQUENCE 522 AA; 59360 MW; AAF305808E9EA75E CRC64;

Query Match      77.6%; Score 45; DB 13; Length 522;
Best Local Similarity 77.8%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GPRPFHCRO 9
DB      136 GERPFHCNO 144

RESULT 7
ID      0922Z2      PRELIMINARY:      PRT:      533 AA.
AC      0922Z2;
DT      01-MAY-1999 (TREMBlrel. 10, Created)
DT      01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      EOS PROTEIN.
GN      ZNFN1A4 OR EOS.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ICR.
RX      MEDLINE=99232954; PubMed=10218586;
RA      Homma Y., Kiyosawa H., Mori T., Oguri A., Nikaide T., Kanazawa K.,
RA      Tojo M., Takeda J., Tanno Y., Yokoya S., Kawabata I., Ikeda H.,
RA      Wanka A.;
RT      "Eos: a novel member of the Ikaros gene family expressed predominantly
RT      in the developing nervous system.";
RL      FEBS Lett. 447:76-80(1999).
DR      EMBL; AB017615; BAA36213.1; -.
DR      HSSP; P15822; 1BBO.
DR      MGD; MGI:1343139; Znf1a4.
DR      InterPro; IPR000822; Znf_C2H2.
DR      Pfam; PF00096; zf-C2H2; 6.
DR      SMART; SM00355; Znf_C2H2; 6.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.

DR      PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW      DNA-binding; Metal-binding; Zinc-finger.
SQ      SEQUENCE 533 AA; 58167 MW; 7A5FF32C6FFDC372 CRC64;

Query Match      77.6%; Score 45; DB 11; Length 533;
Best Local Similarity 77.8%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GPRPFHCRO 9
DB      130 GERPFHCNO 138

RESULT 8
ID      096UP3      PRELIMINARY:      PRT:      545 AA.
AC      096UP3;
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      KIAA1782 PROTEIN (FRAGMENT).
GN      KIAA1782.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=BRAIN;
RX      MEDLINE=21245130; PubMed=11347906;
RA      Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. XX.
RT      The complete sequences of 100 new cDNA clones from brain which code
RT      for large proteins in vitro.";
RL      DNA Res. 8:85-95(2001).
DR      EMBL; AB058685; BBA47411.1; -.
FT      NON-TER      1
SQ      SEQUENCE 545 AA; 59742 MW; 7A8539E5B8FB9BD84 CRC64;

Query Match      77.6%; Score 45; DB 4; Length 545;
Best Local Similarity 77.8%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GPRPFHCRO 9
DB      143 GERPFHCNO 151

RESULT 9
ID      090W82      PRELIMINARY:      PRT:      563 AA.
AC      090W82;
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      LYMPHOID TRANSCRIPTION FACTOR.
GN      HELIOS.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RPL; TISSUE=THYMUS, AND BURSA OF FABRICIUS;
RA      Kohonen P.J., Nera K.P., Liippo J.P., Lassila O.;
RT      "Avian Helios encodes a novel exon in the Ikaros family.";
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ289886; CAC59948.1; -.
SQ      SEQUENCE 563 AA; 63398 MW; 8C1B8ADB34629103 CRC64;

```

Query Match 77.6%; Score 45; DB 13; Length 563;  
 Best Local Similarity 77.8%; Pred. No. 3.3;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 137 GPRPHCRQ 145

RESULT 10  
 ID 096S24 PRELIMINARY; PRT; 725 AA.  
 AC 096S24;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE CDNA FL114549 FIS, CLONE NT2RM2001670, WEAKLY SIMILAR TO ZINC  
 DE FINGER PROTEIN 29.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunaga H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
 RT "MEDO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027455; BAB55124.1;  
 SQ SEQUENCE 725 AA; 80386 MW; 046163DA13669F12 CRC64;

Query Match 77.6%; Score 45; DB 4; Length 725;  
 Best Local Similarity 77.8%; Pred. No. 4.2;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 547 GPRPHCRQ 555

RESULT 11  
 ID 090WJ6 PRELIMINARY; PRT; 428 AA.  
 AC 090WJ6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE IKAROS.  
 OS Seriola quinqueradiata (Five-ray yellowtail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;  
 OC Carangidae; Seriola.  
 OX NCBI\_TaxID=8161;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Sakai M., Okamoto K., Kono T.;  
 RT "Characterization of yellowtail ikaros, a gene necessary for  
 RT differentiation of the immune system."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB060640; BAB43903.1;  
 SQ SEQUENCE 428 AA; 47492 MW; 8F4CBF405A330767 CRC64;

Query Match 75.9%; Score 44; DB 13; Length 428;  
 Best Local Similarity 77.8%; Pred. No. 3.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 52 GPRPHCRQ 60

RESULT 12  
 ID 002056 PRELIMINARY; PRT; 441 AA.  
 AC 002056;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE D HORDEIN (FRAGMENT).  
 GN HOR3.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HIPROLY; TISSUE=ENDOSPERM;  
 RX MEDLINE=92353095; PubMed=1643086;  
 RA Hallford N.G., Tatham A.S., Sui E., Daroda L., Dreyer T., Shewry P.R.;  
 RT "Identification of a novel beta-turn-rich repeat motif in the D  
 RT hordeins of barley."  
 RL Biochim. Biophys. Acta 1122:118-122(1992).  
 DR EMBL; X68072; CAA48209.1;  
 DR InterPro: IPR001419; Glutelin.  
 DR PRINTS; PR00210; GLUTENIN.  
 FT NON\_TER 1  
 SQ SEQUENCE 441 AA; 45994 MW; 60ED394BC756FF44 CRC64;

Query Match 72.4%; Score 42; DB 10; Length 441;  
 Best Local Similarity 66.7%; Pred. No. 9.1;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 262 GPRPHCRQ 270

RESULT 13  
 ID 09ULM2 PRELIMINARY; PRT; 553 AA.  
 AC 09ULM2;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE KIAA1198 PROTEIN (FRAGMENT).  
 GN KIAA1198.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=20039619; PubMed=10574462;  
 RA Nagase T., Ishikawa K., Kikuno R., Hirokawa M., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:337-345(1999).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL; AB033024; BAA8512.1;  
 DR HSSP; P25490; IUBD.  
 DR InterPro: IPR001909; KRAB.  
 DR InterPro: IPR000822; ZnF-C2H2.  
 DR Pfam: PF01352; KRAB; 1.  
 DR Pfam: PF00096; zf-C2H2; 13.  
 DR PRINTS; PR00048; ZINCINGER.

DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; ZnF\_C2H2; 13.  
DR PROSITE; PS50805; KRAB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 13.  
KW DNA-binding; Metal-binding; Nuclear protein; zinc-finger.  
FT NON\_TER  
SQ SEQUENCE 553 AA; 63668 MW; 2B4679D72084CD55 CRC64;

Query Match 72.4%; Score 42; DB 4; Length 553;  
Best Local Similarity 77.8%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
1 1111111  
DB 522 GERPFQCRQ 530

## RESULT 14

O96CP9 PRELIMINARY; PRT; 654 AA.  
AC O96CP9.  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
DE SIMILAR TO EAF TRANSCRIPTION FACTOR 1.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC014068; AAHL4068.1;  
SQ SEQUENCE 654 AA; 70361 MW; C69CC15CE9743298 CRC64;

Query Match 72.4%; Score 42; DB 4; Length 654;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
1111111  
DB 459 GPRPFQCAQ 467

## RESULT 15

O40054 PRELIMINARY; PRT; 707 AA.  
AC O40054.  
DT 01-NOV-1996 (TRENBLREL. 01, Created)  
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
DE D HORDEIN PRECURSOR.  
GN HOR3.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HARUNA N10; TISSUE-ENDOSPERM;  
RA Hirota N., Kuroda H., Ito K.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D82941; BAA1642.1;  
DR InterPro; IPR003612; AAI.  
DR InterPro; IPR001768; Cereal\_1tryp\_aml\_1nh.  
DR InterPro; IPR001419; Glutenin.  
DR Pfam; PF00234; tryp\_alpha\_aml1; 1.

DR PRINTS; PR00210; GLUTENIN.  
DR SMART; SM00459; AAI; 1.  
KW SIGNAL  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 707 D HORDEIN.  
SQ SEQUENCE 707 AA; 75108 MW; EBB873AF035D5789 CRC64;

Query Match 72.4%; Score 42; DB 10; Length 707;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
1 1111111  
DB 528 GKOPFHCOQ 536

Search completed: May 24, 2002, 17:02:51  
Job time: 881 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:03:26 ; Search time 31.45 Seconds

(without alignments)  
11.080 Million cell updates/sec

Title: US-09-730-379E-3

Sequence: 1 GPRFHCRO 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	525	1 HRG_HUMAN	P04196 homo sapien
2	45	77.6	526	1 HELI_HUMAN	Q9UK87 homo sapien
3	45	77.6	526	1 HELI_MOUSE	P81183 mus musculu
4	41	70.7	703	1 EGFR_CHICK	P13387 gallus gall
5	40	69.0	1845	1 2236_HUMAN	Q9U136 homo sapien
6	39	67.2	271	1 RL5_DUNSA	O22608 dunalielella
7	39	67.2	422	1 NSMA_RAT	O96t04 ratius norv
8	39	67.2	533	1 HMHL_DUGTI	Q00400 dugesia tig
9	39	67.2	543	1 ZN08_HUMAN	P17098 homo sapien
10	39	67.2	697	1 KPCG_HUMAN	P05129 homo sapien
11	39	67.2	697	1 KPCG_MOUSE	P05697 mus musculu
12	39	67.2	697	1 KPCG_RABIT	P10829 coryctolagus
13	39	67.2	1829	1 Y296_HUMAN	O15015 homo sapien
14	38	65.5	540	1 Z136_HUMAN	P52737 homo sapien
15	37	63.8	510	1 CEF2_DROME	P20308 drosophila
16	37	63.8	518	1 IKAR_CHICK	O42410 gallus gall
17	37	63.8	519	1 IKAR_HUMAN	O13422 homo sapien
18	37	63.8	645	1 ZF93_MOUSE	O61116 mus musculu
19	37	63.8	707	1 ZF60_MOUSE	P16374 mus musculu
20	37	63.8	751	1 Z337_HUMAN	Q9Y3M9 homo sapien
21	37	63.8	1058	1 POL3_DROME	P04323 drosophila
22	37	63.8	1075	1 NEFC3_HUMAN	O12968 homo sapien
23	37	63.8	1075	1 NEFC3_MOUSE	P97305 mus musculu
24	37	63.8	1350	1 XE1N_XENLA	P08045 xenopus lae
25	36	62.1	386	1 SUCC_COXBU	P53592 coxiella bu
26	36	62.1	393	1 ZF46_MOUSE	Q03309 mus musculu
27	36	62.1	406	1 ZF59_MOUSE	P16733 mus musculu
28	36	62.1	485	1 HUNB_CLOAL	O96785 clogmia alb
29	36	62.1	522	1 IKAR_ONCMY	O13089 oncorhynch
30	36	62.1	532	1 ZN20_HUMAN	P17024 homo sapien
31	36	62.1	626	1 Z189_HUMAN	O75820 homo sapien
32	36	62.1	738	1 ZN84_HUMAN	P51523 homo sapien
33	36	62.1	895	1 Z281_HUMAN	Q9Y2X9 homo sapien

34	36	62.1	1355	1 SALM_DROME	P39770 drosophila
35	36	62.1	1402	1 SALM_DROVI	P39806 drosophila
36	36	62.1	1593	1 AT12_HUMAN	P58397 homo sapien
37	35	60.3	159	1 YKCA_CAEEL	P41995 caenorhabdl
38	35	60.3	188	1 ADML_CANFA	O77559 canis famli
39	35	60.3	506	1 Z157_HUMAN	P51786 homo sapien
40	35	60.3	523	1 FZD2_CHICK	O91A06 gallus gall
41	35	60.3	706	1 BCL6_HUMAN	P41182 homo sapien
42	35	60.3	706	1 Z151_CHICK	O90625 gallus gall
43	35	60.3	727	1 BCL6_MOUSE	P41183 mus musculu
44	35	60.3	727	1 YM29_YEAST	Q03792 saccharomyc
45	35	60.3	794	1 Z151_MOUSE	O60821 mus musculu

## ALIGNMENTS

RESULT 1

ID	HRG_HUMAN	STANDARD:	PRT:	525 AA.
AC	P04196:			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).			
GN	HRG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=66216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";			
RT	Biochemistry 25:2220-2225(1986).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.;			
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";			
RT	Genomics 19:195-197(1994).			
RL	[4]			
RP	SEQUENCE OF 19-27.			
RA	TISSUE=Plasma;			
RC	MEDLINE=93092937; PubMed=1459097;			
RX	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquall C., Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B., Hochstrasser D.F.;			
RA	"Plasma protein map: an update by microsequencing.";			
RT	Electrophoresis 13:707-714(1992).			
RL	CC			
CC	- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS			
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE			
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, 'THROMBOSPONIN'			
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS			
CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN			
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD			
CC	COAGULATION CASCADE.			
CC	- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS			
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-			
CC	RESIDUE SEQUENCE (GHHHP, CONSENSUS) FORM A HISTIDINE-RICH REGION.			
CC	- SIMILARITY: CONTAINS 2 CYSTATHIOLINE DOMAINS.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M13149; AAA52694.1; -  
 DR EMBL: AB005803; BAA21613.1; -  
 DR EMBL: Z17218; CAA78925.1; -  
 DR PIR: A01287; KGHUGH.  
 DR SWISS-2DPAGE; P04196; HUMAN.  
 DR MIM: 142640; -  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; cystatin; 1.  
 DR SMART: SM00043; cy; 2.  
 KW Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
 FT DOMAIN 19 136 CYSTATIN-LIKE 1.  
 FT DOMAIN 137 254 CYSTATIN-LIKE 2.  
 FT DOMAIN 276 321 PRO-RICH.  
 FT DOMAIN 350 497 PRO/HIS-RICH.  
 FT DISULFID 24 504 BY SIMILARITY.  
 FT DISULFID 78 89 BY SIMILARITY.  
 FT DISULFID 105 126 BY SIMILARITY.  
 FT DISULFID 203 417 BY SIMILARITY.  
 FT DISULFID 218 241 BY SIMILARITY.  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 100.0%; Score 58; DB 1; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 0.0049;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 446 GPRPHCRQ 454

RESULT 2  
 HELI\_HUMAN STANDARD; PRT; 526 AA.  
 AC 09UKS7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zinc finger protein Helios.  
 GN ZNFN1A2 OR HELIOS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20009537; PubMed=10541817;  
 RA Hosokawa Y., Maeda Y., Seto M.;  
 RT "Human Helios, an Ikaros-related zinc finger DNA binding protein: cDNA  
 RT cloning and tissue expression pattern.";  
 RL Immunogenetics 50:106-108(1999).  
 CC - FUNCTION: ASSOCIATES WITH IKAROS AT CENTROMERIC HETEROCHROMATIN.  
 CC - SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC - SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF130863; AAP09441.1; -  
 DR HSSP: P15822; IBBO.  
 DR MIM: 606234; -  
 DR InterPro: IPR000822; ZnF-C2H2.  
 DR Pfam: PF00096; zf-C2H2; 5.  
 DR SMART: SM00355; ZnF\_C2H2; 6.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Transcription regulation; Activator; zinc-finger; Metal-binding;  
 KW DNA-binding; Nuclear protein; Repeat.  
 FT ZN\_FING 112 134 C2H2-TYPE.  
 FT ZN\_FING 140 162 C2H2-TYPE.  
 FT ZN\_FING 168 190 C2H2-TYPE.  
 FT ZN\_FING 196 219 C2H2-TYPE.  
 FT ZN\_FING 471 493 C2H2-TYPE.  
 FT ZN\_FING 499 523 C2H2-TYPE.  
 SQ SEQUENCE 526 AA; 59558 MW; A834CB0E0E096647 CRC64;

Query Match 77.6%; Score 45; DB 1; Length 526;  
 Best Local Similarity 77.8%; Pred. No. 0.92;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 136 GPRPHCRQ 144

RESULT 3  
 HELI\_MOUSE STANDARD; PRT; 526 AA.  
 AC P81183;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zinc finger protein Helios.  
 GN ZNFN1A2 OR HELIOS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 246-267 AND 289-306.  
 RC STRAIN=BA1B/C; TISSUE=Thymus;  
 RX MEDLINE=98180961; PubMed=9512513;  
 RA Hahn K., Cobb B.S., McCarty A.S., Brown K.E., Kling C.A., Lee R.,  
 RA Akashi K., Weissman I.L., Fisher A.G., Smale S.T.;  
 RT "Helios, a T cell-restricted Ikaros family member that quantitatively  
 RT associates with Ikaros at centromeric heterochromatin.";  
 RL Genes Dev. 12:782-796(1998).  
 CC - FUNCTION: ASSOCIATES WITH IKAROS AT CENTROMERIC HETEROCHROMATIN.  
 CC - SUBCELLULAR LOCATION: Nuclear.  
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; A AND B (SHOWN HERE); ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC - TISSUE SPECIFICITY: RESTRICTED TO THE T-CELL LINEAGE. ABUNDANT IN  
 CC THYMUS. LOW EXPRESSION IN BONE MARROW AND BRAIN AND  
 CC NO DETECTABLE EXPRESSION IN SPLEEN, LIVER, KIDNEY OR MUSCLE.  
 CC - SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF044257; AAC00513.1; -  
 DR HSSP: P15822; IBBO.



```

DR MGD; MGI:1342541; Znf1a2.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINCINGER.
DR SMART: SM00355; Znf_C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 4.
DR Transcription regulation: Activator; Zinc-finger; Metal-binding;
KM DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 112 219 ZINC-FINGERS I.
FT ZN_FING 112 134 C2H2-TYPE.
FT ZN_FING 140 162 C2H2-TYPE.
FT ZN_FING 168 190 C2H2-TYPE.
FT ZN_FING 196 219 C2H2-TYPE.
FT DOMAIN 471 523 ZINC-FINGERS II.
FT ZN_FING 471 493 C2H2-TYPE.
FT ZN_FING 499 523 C2H2-TYPE.
FT VARSPLIC 111 136 MISSING (IN ISOFORM A).
SQ SEQUENCE 526 AA; 59388 MW; EFD827FE8916108E CRC64;

Query Match 77.6%; Score 45; DB 1; Length 526;
Best Local Similarity 77.8%; Pred. No. 0.92;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9
   1 11111 1
Db 136 GERPHCRQ 144

RESULT 4
EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
CC CELL PROLIFERATION.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL: M20386; AAA48760.1; -.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_Ldomain; 2.
DR SMART: SM00261; Fu_4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A650 CRC64;

Query Match 70.7%; Score 41; DB 1; Length 703;
Best Local Similarity 85.7%; Pred. No. 6;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPHCR 7
   1 11111
Db 524 GPRPHCR 530

RESULT 5
2236_HUMAN STANDARD; PRT; 1845 AA.
AC Q9UL36; Q9UL37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 236.
GN ZNF236.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney;
RX MEDLINE=99389731; PubMed=10458916;
RA Holmes D.I., Wahab N.A., Mason R.M.;
RT "Cloning and characterization of ZNF236, a glucose-regulated kruppel-
RT like zinc-finger gene mapping to human chromosome 19q22-q23.";
RL Genomics 60:105-109(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A AND B (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UNIDIGITOUS. EXPRESSION LEVELS ARE HIGHEST IN
CC SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND
CC PLACENTA, AND LOWEST IN KIDNEY, LIVER, AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
```

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF085244; AAD53329.1; -  
 DR EMBL: AF085243; AAD53328.1; -  
 DR HSP: P07248; IARD.  
 DR MIM: 604760; -  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00096; Zf-C2H2; 30.  
 DR SMART: SM00355; Znf-C2H2; 30.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 30.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 30.  
 DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat; Alternative splicing.  
 FT DOMAIN 37 1801 ZINC\_FINGERS.  
 FT ZN\_FING 37 59 C2H2-TYPE.  
 FT ZN\_FING 66 88 C2H2-TYPE.  
 FT ZN\_FING 93 115 C2H2-TYPE.  
 FT ZN\_FING 121 143 C2H2-TYPE.  
 FT ZN\_FING 133 175 C2H2-TYPE.  
 FT ZN\_FING 197 219 C2H2-TYPE.  
 FT ZN\_FING 225 247 C2H2-TYPE.  
 FT ZN\_FING 253 276 C2H2-TYPE.  
 FT ZN\_FING 285 308 C2H2-TYPE.  
 FT ZN\_FING 482 504 C2H2-TYPE.  
 FT ZN\_FING 510 532 C2H2-TYPE.  
 FT ZN\_FING 538 560 C2H2-TYPE.  
 FT ZN\_FING 566 588 C2H2-TYPE.  
 FT ZN\_FING 657 679 C2H2-TYPE.  
 FT ZN\_FING 685 707 C2H2-TYPE.  
 FT ZN\_FING 713 735 C2H2-TYPE.  
 FT ZN\_FING 741 763 C2H2-TYPE.  
 FT ZN\_FING 967 989 C2H2-TYPE.  
 FT ZN\_FING 995 1017 C2H2-TYPE.  
 FT ZN\_FING 1023 1045 C2H2-TYPE.  
 FT ZN\_FING 1051 1073 C2H2-TYPE.  
 FT ZN\_FING 1166 1189 C2H2-TYPE.  
 FT ZN\_FING 1167 1189 C2H2-TYPE.  
 FT ZN\_FING 1195 1217 C2H2-TYPE.  
 FT ZN\_FING 1223 1245 C2H2-TYPE.  
 FT ZN\_FING 1251 1273 C2H2-TYPE.  
 FT ZN\_FING 1657 1680 C2H2-TYPE.  
 FT ZN\_FING 1686 1708 C2H2-TYPE.  
 FT ZN\_FING 1722 1744 C2H2-TYPE.  
 FT ZN\_FING 1750 1772 C2H2-TYPE.  
 FT ZN\_FING 1778 1801 C2H2-TYPE.  
 FT VARSPLIC 1530 1558 ELNWTSGSLPTPTSPSAISTONLWSS -> GSRSVQHS  
 FT VARSPLIC 1559 1845 VGPOEGSGVEALYLENNSDKT (IN ISOCORM A).  
 FT SEQUENCE 1845 AA; 203659 MW; 2879EA91D0C6D3D8 CRC64;  
 SQ

Query Match 69.0%; Score 40; DB 1; Length 1845;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPRPFC 7  
 Db 1746 GERPFC 1752

RESULT 6  
 RL5\_DUNSA STANDARD; PRT; 271 AA.  
 AC 022608;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 60S ribosomal protein L5.  
 GN RP5 OR DSRP1.  
 OS Dunaliella salina.  
 OC Eukaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales;  
 OC Dunaliellaceae; Dunaliella.  
 NCBI\_TaxID=3046;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA KO J.H., Lee S.H.;  
 RT "A cDNA encodes a protein sequence homologous to the eukaryotic  
 RT ribosomal 5S RNA-binding protein from Dunaliella salina";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF028833; AAB84056.1; -  
 DR InterPro: IPR001149; Ribosomal\_L18p.  
 DR Pfam: PF00861; Ribosomal\_L18p; 1.  
 DR PRINTS: PR00058; RIBOSOMAL15.  
 DR ProDom: PD001394; Ribosomal\_L18p; 1.  
 KW Ribosomal protein; rRNA-binding  
 SQ SEQUENCE 271 AA; 31218 MW; 482P5383D3B57E CRC64;  
 QY 1 GPRPFC 7  
 Db 138 GPRPFC 144

Query Match 67.2%; Score 39; DB 1; Length 271;  
 Best Local Similarity 71.4%; Pred. No. 5,6;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPRPFC 7  
 Db 138 GPRPFC 144

RESULT 7  
 ID NSMA\_RAT STANDARD; PRT; 422 AA.  
 AC Q9ET64;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Sphingomyelin phosphodiesterase 2 (EC 3.1.4.12) (Neutral  
 DE sphingomyelinase) (nsnase) (N-Snase) (Lyso-platelet activating factor-  
 DE phospholipase C) (Lyso-PAF-PLC).  
 GN SMPD2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.; CHARACTERIZATION, AND MUTAGENESIS.  
 RC SPRAIN-Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=20282884; Pubmed=10832103;  
 RA Mizutani Y., Tamiya-Koizumi K., Irie F., Hirabayashi Y., Miwa M.,  
 RA Yoshida S.;  
 RT "Cloning and expression of rat neutral sphingomyelinase:  
 RT enzymological characterization and identification of essential  
 RT histidine residues";  
 RL Biochim. Biophys. Acta 1485:236-246(2000).  
 CC -1- FUNCTION: Converts sphingomyelin to ceramide. Hydrolyze 1-acyl-2-  
 CC lyso-sn-glycero-3-phosphocholine (lyso-PC) and 1-O-alkyl-2-lyso-  
 CC sn-glycero-3-phosphocholine (lyso-platelet activating factor). The  
 CC physiological substrate seems to be lyso-PAF.  
 CC -1- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +  
 CC choline phosphate.

```

CC -1- COFACTOR: Magnesium.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AB047002; BAB08219.1; -
KW Hydrolase; Transmembrane; Magnesium.
FT TRANSMEM 325 345 POTENTIAL.
FT METAL 354 374
FT SITE 180 180 IMPORTANT FOR SUBSTRATE RECOGNITION (BY
FT ACT SITE 272 272 SIMILARITY).
FT MUTAGEN 136 136 GENERAL BASE (PROBABLE).
FT MUTAGEN 151 151 H->A: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 151 151 H->Y: REDUCED ACTIVITY.
FT MUTAGEN 272 272 H->Y: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 272 272 H->A: COMPLETE LOSS OF ACTIVITY.
SQ SEQUENCE 422 AA; 47644 MW; 109A5133A056AAFI CRC64;

Query Match 67.2%; Score 39; DB 1; Length 422;
Best Local Similarity 75.0%; Pred. No. 8.4;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 PRPFHCRO 9
Db 406 PRPFHCRO 413

RESULT 8
HMH1_DUGHI STANDARD; PRT; 533 AA.
ID HMH1_DUGHI
AC 000400;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Homeobox protein DTH-1.
GN DTH-1.
OS Dugesia tigrina (Planarian).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhabdiorhiza; Sericata; Tricladida; Paludicola; Dugesidae; Girardia.
OX NCBI_TaxID=6162;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93387216; PubMed=8104142;
RA Garcia-Fernandez J., Baguna J., Salo E.;
RT "Genomic organization and expression of the planarian homeobox genes
RT Dth-1 and Dth-2."
RL Development 118:241-253(1993).
RN [2]
RP SEQUENCE OF 83-533 FROM N.A.
RX MEDLINE=91334461; PubMed=1714599;
RA Salo E., Garcia-Fernandez J., Baguna J.;
RT "Planarian homeobox genes: cloning, sequence analysis, and
RT expression."
RL Proc. Natl. Acad. Sci. U.S.A. 88:7338-7342(1991).
CC -1- FUNCTION: THIS PROTEIN MIGHT BE INVOLVED IN DETERMINATION AND/OR
CC DIFFERENTIATION OF NERVE CELLS IN THE CONTINUOUS REPLACEMENT OF
CC NEURONS IN THE CEPHALIC REGION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: INTESTINE AND UNIDENTIFIED PERIPHERAL
CC PARENCHYMAL CELLS. SLIGHTLY HIGHER LEVELS IN THE CEPHALIC REGION
CC COMPARED TO OTHER BODY REGIONS.
CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X69203; CAA49141.1; -
DR EMBL: X69200; CAA49141.1; JOINED.
DR EMBL: X69201; CAA49141.1; JOINED.
DR EMBL: X56499; CAA39854.1; -
DR PIR: A4151; A4151.
DR PIR: S33701; S33701.
DR HSSP: P22808; INK3.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PRO0024; HOMEBOX.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS0071; HOMEBOX_2; 1.
KV Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 85 90
FT DOMAIN 100 270 PRO-RICH.
FT DOMAIN 283 371 ASN/SER-RICH.
FT DOMAIN 342 351 ASN/GLU-RICH (ACIDIC).
FT DNA BIND 377 436 HOMEBOX.
FT DOMAIN 510 533 ASN/SER-RICH.
SQ SEQUENCE 533 AA; 60275 MW; 2AC6521DBEACAE CRC64;

Query Match 67.2%; Score 39; DB 1; Length 533;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRFHCRO 9
Db 228 GPRFHCRO 236

RESULT 9
ZNO8_HUMAN STANDARD; PRT; 543 AA.
ID ZNO8_HUMAN
AC P17098;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 8 (Zinc finger protein Hf.18) (Fragment).
GN ZNF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90169993; PubMed=2106481;
RA Lania L., Donli E., Pannuti A., Pascucci A., Penque G.,
RA Feliciello I., la Mantia G., Ianfrancone L., Pelicci P.-G.;
RT "cDNA isolation, expression analysis, and chromosomal localization of
RT two human zinc finger genes."
RL Genomics 6:333-340(1990).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: UNIDIRECTIONALLY PRESENT IN MANY HUMAN CELL LINES
CC OF DIFFERENT EMBRYOLOGICAL DERIVATION.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M29581; AAA61314.1; -  
 DR PIR: B34612; B34612.  
 DR HSSP: P08047; 1SP2.  
 DR MIM: 194532; -  
 DR InterPro: IPR001909; KRAB.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF01352; KRAB; 1.  
 DR Pfam: PF00096; zf-C2H2; 7.  
 DR SMART: SM00349; KRAB; 1.  
 DR SMART: SM00353; Znf-C2H2; 7.  
 DR PROSITE: PS50805; KRAB; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KM Nuclear protein; Repeat.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 64 KRAB.  
 FT ZINC\_FINGERS.  
 FT ZN\_FING 225 457 C2H2-TYPE.  
 FT ZN\_FING 225 247 C2H2-TYPE.  
 FT ZN\_FING 253 275 C2H2-TYPE.  
 FT ZN\_FING 281 303 C2H2-TYPE.  
 FT ZN\_FING 309 331 C2H2-TYPE.  
 FT ZN\_FING 337 359 C2H2-TYPE.  
 FT ZN\_FING 365 387 C2H2-TYPE.  
 FT ZN\_FING 435 457 C2H2-TYPE.  
 SQ SEQUENCE 543 AA; 61772 MW; ADD987504ECAC019 CRC64;

Query Match 67.2%; Score 39; DB 1; Length 543;  
 Best Local Similarity 66.7%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
 DB 389 GERPECRQ 397

RESULT 10  
 KRCG\_HUMAN  
 ID KRCG\_HUMAN STANDARD; PRT; 697 AA.  
 AC P05129;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein kinase C, gamma type (EC 2.7.1.-) (PKC-gamma).  
 GN PKC-G OR PKCG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cui W.C., Yu L., Chu Y.Y., Wang J., Zheng L.H., Zhou G.J., Zhao S.Y.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-318 FROM N.A.  
 RC TISSUE-Brain;  
 RX MEDLINE=86289426; PubMed=3755548;  
 RA Cousens L., Parker P.J., Rhee L., Yang-Feng T.L., Chen E.,  
 RA Waterfield M.D., Francke U., Ullrich A.;  
 RT "Multiple, distinct forms of bovine and human protein kinase C  
 RT suggest diversity in cellular signaling pathways.";  
 RL Science 233:859-866(1986).  
 RN [3]  
 RP SEQUENCE OF 162-697 FROM N.A.  
 RC TISSUE-Hippocampus;  
 RX MEDLINE=93387312; PubMed=8375396;  
 RA Kochs G., Meyer D., Hug H., Marne D., Sarre T.F.;  
 RT "Activation and substrate specificity of the human protein kinase C

RT alpha and zeta isoenzymes.";  
 RL Eur. J. Biochem. 216:597-606(1993).  
 RN [4]  
 RP VARIANTS C-141; O-415; D-523 AND S-659.  
 RX MEDLINE=98213587; PubMed=9545390;  
 RA Al-Maghtheh M., Vilhena E.N., Inglehearn C.F., Moore T., Bird A.C.,  
 RA Bhattacharya S.S.;  
 RT "Segregation of a PKCG mutation in two RP11 families.";  
 RL Am. J. Hum. Genet. 62:1248-1252(1998).  
 RN [5]  
 RP SHOWS THAT THE VARIANTS ARE NOT A CAUSE OF RP11.  
 RX MEDLINE=99375047; PubMed=10441600;  
 RA Dryja T.P., McEvoy J., McGee T.L., Berson E.L.;  
 RT "No mutations in the coding region of the PRKG gene in three families  
 RT with retinitis pigmentosa linked to the RP11 locus on chromosome  
 RT 19q.";  
 RL Am. J. Hum. Genet. 65:926-928(1999).  
 CC -1- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,  
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.  
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC PKC SUBFAMILY.  
 CC -1- DATABASE: NAME-Mutations of the PRKG gene;  
 CC NOTE-Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/prkgmut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF345987; AAK13533.1; -  
 DR EMBL: M13977; AAA60102.1; ALT\_TERM.  
 DR PIR: D24664; D24664.  
 DR HSSP: P05697; ITBN.  
 DR MIM: 176980; -  
 DR InterPro: IPR000008; C2.  
 DR InterPro: IPR002219; DAG\_PE-bind.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR000961; pkinase\_C.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00168; C2; 1.  
 DR Pfam: PF00130; DAG\_PE-bind; 2.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00433; pkinase\_C; 1.  
 DR PRINTS: PR00360; C2DOMAIN.  
 DR PRINTS: PR00008; DAGPEDOMAIN.  
 DR SMART: SM00109; C1; 2.  
 DR SMART: SM00239; C2; 1.  
 DR SMART: SM00133; S\_TK\_X; 1.  
 DR SMART: SM00220; S\_TK\_X; 1.  
 DR PROSITE: PS00499; C2\_DOMAIN\_1; 1.  
 DR PROSITE: PS00499; C2\_DOMAIN\_2; 1.  
 DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
 DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_2; 2.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Calcium-binding; Repeat; ATP-binding; Transferase;  
 KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc;  
 KW Phosphorylation; Polymorphism.  
 FT DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.  
 FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.  
 FT DOMAIN 170 260 C2 DOMAIN.

```

FT DOMAIN 351 614 PROTEIN KINASE.
FT NP_BIND 357 365 APP (BY SIMILARITY).
FT BINDING 380 380 APP (BY SIMILARITY).
FT ACT_SITE 480 480 BY SIMILARITY.
FT MOD_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT VARIANT 141 141 R->C.
FT VARIANT 415 415 /FTID-VAR_008755.
FT VARIANT 523 523 H->O.
FT VARIANT 523 523 /FTID-VAR_008756.
FT VARIANT 659 659 A->D.
FT VARIANT 659 659 /FTID-VAR_008757.
FT VARIANT 659 659 R->S.
FT SEQUENCE 697 AA: 78447 MW: 3F91B5BEF713C41 CRC64:

Query Match 67.2% Score 39, DB 1, Length 697,
Best Local Similarity 66.7% Pred. No. 13,
Matches 6, Conservative 1, Mismatches 2, Indels 0, Gaps 0;

QY 1 GRRPFCRQ 9
Db 14 GRRPFCRQ 22

RESULT 11
KRCG_MOUSE
ID KRCG_MOUSE STANDARD; PRT: 697 AA.
AC P05697;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein kinase C, gamma type (EC 2.7.1.1) (PKC-gamma).
GN PKCG OR PKCG OR PKCC.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN 11
RP SEQUENCE FROM N.A.
RC SPECIES-Rat; TISSUE-Brain;
RX MEDLINE=86272097; PubMed=3755379;
RA Ono Y., Fujii T., Igarashi K., Kikkawa U., Ogita K., Nishizuka Y.;
RT "Nucleotide sequences of cDNAs for alpha and gamma subspecies of rat
RT brain protein kinase C."
RL Nucleic Acids Res. 16:5199-5200(1988).
RN 12
RP SEQUENCE FROM N.A.
RC SPECIES-Rat;
RX MEDLINE=86272097; PubMed=3755379;
RA Knopf J.L., Lee M.-H., Sultzman L.A., Kriz R.W., Loomis C.R.,
RT Hewick R.M., Bell R.M.;
RT "Cloning and expression of multiple protein kinase C cDNAs."
RL Cell 46:491-502(1986).
RN 13
RP SEQUENCE OF 1-56 FROM N.A.
RC SPECIES-Rat;
RX MEDLINE=91060619; PubMed=2246272;
RA Chen K.H., Widen S.G., Wilson S.H., Huang K.P.;
RT "Characterization of the 5'-flanking region of the rat protein kinase
RT C gamma gene."
RL J. Biol. Chem. 265:19961-19965(1990).
RN 14
RP SEQUENCE FROM N.A.
RC SPECIES-Mouse; TISSUE-Brain;
RX MEDLINE=93154595; PubMed=8428669;
RA Bowers B.J., Parham C.L., Sikele J.M., Wehner J.M.;
RT "Isolation and sequence of a mouse brain cDNA coding for protein
RT kinase C-gamma isozyme."
RL Gene 123:263-265(1993).
RN 15

```

```

RP SEQUENCE FROM N.A.
RC SPECIES-Mouse; STRAIN-BALB/C; TISSUE-Brain;
RA Tseng C.P., Verma A.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN 16
RP STRUCTURE BY NMR OF 91-172.
RC SPECIES-Rat;
RX MEDLINE=97419134; PubMed=9271501;
RA Xu R.X., Pawelczyk T., Xia T.-H., Brown S.C.;
RT "NMR structure of a protein kinase C-gamma phorbol-binding domain and
RT study of protein-lipid micelle interactions."
RL Biochemistry 36:10709-10717(1997).
CC -1- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07287; CAA30267.1; -
DR EMBL: M13707; AAA41874.1; -
DR EMBL: M55417; AAA41873.1; -
DR EMBL: X67129; CAA47608.1; -
DR EMBL: L28035; AAA39939.1; -
DR PIR: A05105; KIRTCG.
DR PIR: S29620; S29620.
DR PIR: JN0548; JN0548.
DR PDB: 1TBN; 29-APR-96.
DR PDB: 1TBO; 29-APR-96.
DR MGD: MGI:97597; PKCC.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000961; Pkinase_C.
DR Pfam: PF00168; C2_1.
DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00433; pkinase_C; 1.
DR PRINTS: PR00360; C2DOMAIN.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR SMART: SM00109; C1; 2.
DR SMART: SM00239; C2; 1.
DR SMART: SM00133; S_TK_X; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00499; C2_DOMAIN_1; 1.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR Calcium-binding; Repeat; ATP-binding; Transferase;
DR Serine/threonine-protein kinase; Phorbol-ester binding; zinc;
DR Phosphorylation; 3D-structure.
KW DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 170 260 C2 DOMAIN.
FT DOMAIN 351 614 PROTEIN KINASE.
FT NP_BIND 357 365 ATP (BY SIMILARITY).
FT BINDING 380 380 ATP (BY SIMILARITY).

```

FT ACT\_SITE 480 480 BY SIMILARITY.  
 FT MOD\_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT MOD\_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 SQ SEQUENCE 697 AA; 78357 MW; E6E2F7A3B93042FF CRC64;

Query Match 67.2%; Score 39; DB 1; Length 697;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRQ 9  
 1111 11:  
 Db 14 GPRPLFCRK 22

RESULT 12  
 KPGC\_RABIT  
 ID KPGC\_RABIT STANDARD; PRT; 697 AA.  
 AC P10829;

DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Protein kinase C, gamma type (EC 2.7.1.-) (PKC-gamma) (Delta).  
 GN PKCG.

OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88241036; PubMed=2837282;  
 RA Ohno S., Kawasaki H., Kono Y., Inagaki M., Hidaka H., Suzuki K.;  
 RT "A fourth type of rabbit protein kinase C.";  
 RL Biochemistry 27:2083-2087(1988).

CC -1- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,  
 SERINE- AND THREONINE-SPECIFIC ENZYME.  
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS. A CLASS OF TUMOR PROMOTERS.  
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC PKC SUBFAMILY

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M19338; AAA31449.1; -  
 CC PIR: A28708; KIRBGC.  
 CC HSSP: P05697; ITBN.  
 CC InterPro: IPR000008; C2.  
 CC InterPro: IPR002219; DAG\_PE-bind.  
 CC InterPro: IPR000719; Euk\_pkinase.  
 CC InterPro: IPR000961; Pkinase\_C.  
 CC InterPro: IPR002290; Ser\_thr\_pkinase.  
 CC Pfam: PF00168; C2; 1.  
 CC Pfam: PF00130; DAG\_PE-bind; 2.  
 CC Pfam: PF00069; pkinase; 1.  
 CC PRINTS: PR00360; C2DOMAIN.  
 CC PRINTS: PR00008; DAGPEDOMAIN.  
 CC SMART: SM00109; C1; 2.  
 CC SMART: SM00239; C2; 1.  
 CC SMART: SM00133; S\_TK\_X; 1.  
 CC SMART: SM00220; S\_TKC; 1.  
 CC PROSITE: PS00499; C2\_DOMAIN\_1; 1.  
 CC PROSITE: PS00004; C2\_DOMAIN\_2; 1.

DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
 DR PROSITE: PS50081; DAG\_PE\_BIND\_DOM\_2; 2.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Calcium-binding. Repeat: ATP-binding; Transferase;  
 KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc;  
 KW Phosphorylation.  
 FT DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.  
 FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.  
 FT DOMAIN 170 260 C2 DOMAIN.  
 FT DOMAIN 351 614 PROTEIN KINASE.  
 FT NP\_BIND 357 365 ATP (BY SIMILARITY).  
 FT BINDING 380 380 ATP (BY SIMILARITY).  
 FT ACT\_SITE 480 480 BY SIMILARITY.  
 FT MOD\_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT MOD\_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 SQ SEQUENCE 697 AA; 78371 MW; 925D22221F78E5BF CRC64;

Query Match 67.2%; Score 39; DB 1; Length 697;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRQ 9  
 1111 11:  
 Db 14 GPRPLFCRK 22

RESULT 13  
 Y296\_HUMAN  
 ID Y296\_HUMAN STANDARD; PRT; 1829 AA.  
 AC O15015;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical zinc finger protein KIAA0296.  
 GN KIAA0296.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB002294; BAA20756.1; -  
 CC HSSP: P03001; ITF3.  
 CC InterPro: IPR000822; Znf-C2H2.  
 CC Pfam: PF00096; Zf-C2H2; 31.  
 CC PRINTS: PR00048; ZINC\_FINGER.  
 CC SMART: SM00355; Znf\_C2H2; 30.  
 CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 30.  
 CC PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 38.

KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KM Zinc-finger; Metal-binding; Nuclear protein; Repeat.  
 FT DOMAIN 8 1783 ZINC FINGERS.  
 FT ZN\_FING 8 31 C2H2-TYPE.  
 FT ZN\_FING 48 70 C2H2-TYPE.  
 FT ZN\_FING 75 97 C2H2-TYPE.  
 FT ZN\_FING 239 261 C2H2-TYPE.  
 FT ZN\_FING 266 288 C2H2-TYPE.  
 FT ZN\_FING 294 316 C2H2-TYPE.  
 FT ZN\_FING 374 396 C2H2-TYPE.  
 FT ZN\_FING 401 424 C2H2-TYPE.  
 FT ZN\_FING 465 487 C2H2-TYPE.  
 FT ZN\_FING 492 514 C2H2-TYPE.  
 FT ZN\_FING 575 597 C2H2-TYPE.  
 FT ZN\_FING 617 639 C2H2-TYPE.  
 FT ZN\_FING 644 666 C2H2-TYPE.  
 FT ZN\_FING 821 843 C2H2-TYPE.  
 FT ZN\_FING 848 870 C2H2-TYPE.  
 FT ZN\_FING 958 980 C2H2-TYPE.  
 FT ZN\_FING 981 904 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 1052 1074 C2H2-TYPE.  
 FT ZN\_FING 1101 1101 C2H2-TYPE.  
 FT ZN\_FING 1203 1225 C2H2-TYPE.  
 FT ZN\_FING 1230 1252 C2H2-TYPE.  
 FT ZN\_FING 1258 1280 C2H2-TYPE.  
 FT ZN\_FING 1299 1321 C2H2-TYPE.  
 FT ZN\_FING 1326 1348 C2H2-TYPE.  
 FT ZN\_FING 1364 1386 C2H2-TYPE.  
 FT ZN\_FING 1557 1579 C2H2-TYPE.  
 FT ZN\_FING 1585 1607 C2H2-TYPE.  
 FT ZN\_FING 1677 1699 C2H2-TYPE.  
 FT ZN\_FING 1704 1726 C2H2-TYPE.  
 FT ZN\_FING 1732 1754 C2H2-TYPE.  
 FT ZN\_FING 1761 1783 C2H2-TYPE.  
 SQ SEQUENCE 1829 AA; 200760 MW; B20C482B38684895 CRC64;

Query Match Score 39; DB 1; Length 1829;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 PRPFHCQ 9  
 DB 1049 PRPFHCQ 1056

RESULT 14  
 Z136\_HUMAN STANDARD; PRT; 540 AA.  
 AC P52737;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zinc finger protein 136.  
 GN ZNF136.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Insulinoma;  
 RX MEDLINE=95377390; PubMed=7649249;  
 RA Vissing H., Meyer W.-K., Aagaard L., Tommerup N., Thiesen H.-J.;  
 RT "Repression of transcriptional activity by heterologous KRAB domains  
 present in zinc finger proteins.";  
 RL FEBS Lett. 369:153-157(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION AS A WEAK  
 CC REPRESSOR WHEN ALONE, OR A POTENT ONE WHEN FUSED WITH A  
 CC HETEROLOGOUS PROTEIN CONTAINING A KRAB-B-DOMAIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: SEEMS UBICUITOUS. SEEN IN THE HEART, BRAIN,  
 CC PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, AND PANCREAS.

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U09367; AAC50261.1; -  
 CC HSSP: P08151; ZGLI.  
 CC MIM: 604078; -  
 CC InterPro: IPR001909; KRAB.  
 CC InterPro: IPR000822; Znf-C2H2.  
 CC Pfam: PF01352; KRAB; 1.  
 CC Pfam: PF00096; Zf-C2H2; 13.  
 CC PRINTS: PR00048; ZINCFINGER.  
 CC SMART: SM00349; KRAB; 1.  
 CC SMART: SM00355; Znf-C2H2; 13.  
 CC PROSITE: PS00805; KRAB; 1.  
 CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
 CC PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 14.  
 KM Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
 KW Nuclear protein; Repeat; Repressor.  
 FT DOMAIN 4 90 KRAB.  
 FT ZN\_FING 168 540 ZINC FINGERS.  
 FT ZN\_FING 168 190 C2H2-TYPE.  
 FT ZN\_FING 196 218 C2H2-TYPE.  
 FT ZN\_FING 224 246 C2H2-TYPE.  
 FT ZN\_FING 252 274 C2H2-TYPE.  
 FT ZN\_FING 280 302 C2H2-TYPE.  
 FT ZN\_FING 308 330 C2H2-TYPE.  
 FT ZN\_FING 336 358 C2H2-TYPE.  
 FT ZN\_FING 364 386 C2H2-TYPE.  
 FT ZN\_FING 392 414 C2H2-TYPE.  
 FT ZN\_FING 420 442 C2H2-TYPE.  
 FT ZN\_FING 448 470 C2H2-TYPE.  
 FT ZN\_FING 476 498 C2H2-TYPE.  
 FT ZN\_FING 504 526 C2H2-TYPE.  
 SQ SEQUENCE 540 AA; 62783 MW; 313297AB22F62952 CRC64;

Query Match Score 38; DB 1; Length 540;  
 Best Local Similarity 44.4%; Pred. No. 16;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPHCQ 9  
 DB 500 GOKPYHCKE 508

RESULT 15  
 CF2\_DROME STANDARD; PRT; 510 AA.  
 AC P20385; Q24263; Q9YR41;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chorion transcription factor Cf2, isoforms I and II.  
 GN CF2 OR CG11924.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydriidae; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOPRIMS I AND II).  
 RC STRAIN=OREGON-R; TISSUE=Embryo, and Ovary;  
 RX MEDLINE=93030711; PubMed=1411512;  
 RA Hsu T., Gogos J.A., Kirsh S.A., Kafatos F.C.;

RT "Multiple zinc finger forms resulting from developmentally regulated  
 RT alternative splicing of a transcription factor gene."  
 RL Science 257:1946-1950(1992).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sulten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-U., Andrews-Planck C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beres P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Casley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ideyem C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mialshina N.V., Mobarry C., Morris J., Moshneft A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 [3]  
 RN SEQUENCE OF 248-510 FROM N.A. (ISOFORM II).  
 RP TISSUE=Embryo;  
 RX MEDLINE=91007257; PubMed=2120114;  
 RA Shea M.J., King D.L., Conboy M.J., Mariani B.D., Kafatos F.C.;  
 RT "Proteins that bind to Drosophila chorion cis-regulatory elements: a  
 RT new C2H2 zinc finger protein and a C2C2 steroid receptor-like  
 RT component.";  
 RL Genes Dev. 4:1128-1140(1990).  
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR. ISOFORM II BINDS TO THE  
 CC PROMOTER REGION OF S15 CHORION GENE, WHEREAS ISOFORM II BINDS TO  
 CC CP2 PROMOTER, THUS HAVING A PROBABLE AUTOREGULATORY ROLE.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: I (SHOWN HERE), II AND III (AC  
 CC 001522). ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: ISOFORM I IS FOUND IN EMBRYOS, PUPAE AND ADULT  
 CC SOMATIC TISSUE; ISOFORM II OCCURS IN EMBRYOS, PUPAE, OVARIES AND  
 CC TO A LESSER EXTENT IN ADULT SOMATIC TISSUE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, M97196; AAA28395.1; -

DR EMBL: M97196; AAA28395.1; -;  
 DR EMBL: AE003575; AAF50966.2; -;  
 DR EMBL: X53380; CAA37460.1; -;  
 DR PIR: B35872; B35872.  
 DR PIR: B35872; B35872.  
 DR PIR: B35872; B35872.  
 DR PIR: B35872; B35872.  
 DR HSSP: P08047; ISP2.  
 DR TRANSFAC: T00119; -;  
 DR TRANSFAC: T00120; -;  
 DR FLYBASE: FBgn0000286; CF2.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00096; Zf-C2H2; 7.  
 DR PRINTS: PR00048; ZINCINGER.  
 DR SMART: SM00355; ZNF-C2H2; 7.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
 DR Nuclear protein: DNA-binding; Zinc-finger; Trans-acting factor;  
 KW Transcription regulation; Alternative splicing; Metal-binding; Repeat.  
 KM ZN\_FING 76 97 C2H2-TYPE.  
 FT ZN\_FING 127 138 C2H2-TYPE.  
 FT ZN\_FING 366 388 C2H2-TYPE.  
 FT ZN\_FING 403 423 C2H2-TYPE.  
 FT ZN\_FING 431 451 C2H2-TYPE.  
 FT ZN\_FING 459 479 C2H2-TYPE.  
 FT ZN\_FING 487 508 C2H2-TYPE.  
 FT DOMAIN 21 28 POLY-PRO.  
 FT DOMAIN 238 252 POLY-GLN.  
 FT DOMAIN 263 270 POLY-GLN.  
 FT VARSPIC 430 457 MISSING (IN ISOFORM II).  
 FT CONFLICT 254 254 S -> L (IN REF. 2).  
 SQ SEQUENCE 510 AA; 56724 MW; 4895F851FAEE0AB CRC64;

Query Match 63.8%; Score 37; DB 1; Length 510;  
 Best Local Similarity 71.4%; Pred. No. 22;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GPRPHC 7  
 Db 453 GERPHC 459

Search completed: May 24, 2002, 17:03:27  
 Job time: 862 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 24, 2002, 16:50:11 ; Search time 64.04 Seconds  
(without alignments)  
13.504 Million cell updates/sec

Title: US-09-730-379E-3

Perfect score: 58

Sequence: 1 GPRPHCRQ 9

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: PIR71:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	525	1 KGHUGH	histidine-rich gly
2	45	77.6	431	2 T47528	hypothetical prote
3	42	72.4	707	2 S52390	D-hordein precursor
4	41	70.7	1223	1 TVCHLV	epidermal growth f
5	40	69.0	227	2 S44281	ES43 protein - bar
6	39	67.2	101	2 S72766	hypothetical prote
7	39	67.2	271	2 T08009	probable ribosomal
8	39	67.2	533	2 S33701	homeotic protein D
9	39	67.2	543	2 B34612	zinc finger protei
10	39	67.2	697	1 KIRPGC	protein kinase C (
11	39	67.2	697	1 KIRBGC	protein kinase C (
12	39	67.2	697	2 JN0548	protein kinase C (
13	39	67.2	697	2 D24664	protein kinase C (
14	38	65.5	540	2 B57785	zinc finger protei
15	38	65.5	1891	2 T13594	hypothetical prote
16	38	65.5	1920	2 T13893	gene hindstight pro
17	37	63.8	427	2 B56229	lymphoid transcrip
18	37	63.8	431	2 I59572	Ikabos DNA binding
19	37	63.8	482	2 B36901	chromatin transcrip
20	37	63.8	496	2 T08674	probable finger pr
21	37	63.8	510	2 A36901	chorion transcrip
22	37	63.8	707	2 S68858	finger protein - m
23	37	63.8	1058	1 GNFF17	retrovirus-related
24	37	63.8	1065	2 A57410	transcription fact
25	37	63.8	1075	2 A57377	transcription fact
26	37	63.8	1350	2 S00647	finger protein - A
27	36	62.1	218	2 T27954	hypothetical prote
28	36	62.1	393	2 JN0533	finger protein PML
29	36	62.1	407	2 B39240	finger protein mfg

30	36	62.1	409	2 D83326	probable acyl-CoA
31	36	62.1	564	2 T12489	hypothetical prote
32	36	62.1	624	2 S41688	DNA-binding protei
33	36	62.1	651	2 B37891	finger protein 2,
34	36	62.1	728	2 A48830	probable transcrip
35	36	62.1	895	2 JC7089	zinc finger bindin
36	36	62.1	898	2 T14764	hypothetical prote
37	36	62.1	1028	2 S41749	myosin heavy chain
38	36	62.1	1355	2 S40022	spalt protein - fr
39	36	62.1	1402	2 S42748	finger protein - f
40	35	60.3	16	2 D45193	zinc finger protei
41	35	60.3	77	2 D30502	Ig kappa chain V r
42	35	60.3	159	2 T15298	hypothetical prote
43	35	60.3	240	2 A83462	hypothetical prote
44	35	60.3	302	2 E86267	hypothetical prote
45	35	60.3	312	2 T22918	hypothetical prote

## ALIGNMENTS

## RESULT 1

KGHUGH  
histidine-rich glycoprotein precursor - human  
N:Alternate names: HRG  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000  
C:Accession: A01287; S29659  
R:Koida, T.; Foster, D.; Yoshitake, S.; Davie, E.W.  
Biochemistry 25, 2220-2225, 1986  
A:Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu  
A:Reference number: A01287; MUID:86216149  
A:Accession: A01287  
A:Molecule type: mRNA  
A:Residues: 1-525 <KO1>  
A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514  
R:Hennis, B.; Havelaar, A.; Kluff, C.  
submitted to the EMBL Data Library, October 1991  
A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly  
A:Reference number: S29669  
A:Accession: S29669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 214-247 <HEN>  
A:Cross-references: EMBL:U17218; NID:g32453; PIDN:CAA78925.1; PID:g32454  
C:Comment: Although its physiological function is not yet known, HRG does bind heme,  
din, and the lysine-binding site of plasminogen. On the basis of its homology with Hm  
lood coagulation cascade.  
C:Comment: The amino half of this protein is homologous to the first two cystatin-11k  
could not have inhibitory activity.  
C:Comment: In addition to having a high histidine and proline content, this protein h  
e-rich' region.  
C:Genetics:  
A:Gene: GDB:HRG  
A:Cross-references: GDB:120055; OMIM:142640  
A:Map position: 3q27-3q27  
C:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat  
F:1-18/Domain: signal sequence #status predicted <Sig>  
F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>  
F:140-246/Domain: cystatin homology <CY1>  
F:276-321/Domain: cystatin homology <CY2>  
F:348-437/Region: histidine-rich  
F:351-497/Region: proline-rich  
F:63,125,344,345/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 100.0%; Score 58; DB 1; Length 525;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
DB 446 GPRPFHCQ 454

## RESULT 2

T47528  
hypothetical protein T31B5.20 - Arabidopsis thaliana  
N:Alternate names: phospholipase-like protein; protein F16L2.90  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 05-May-2000  
C:Accession: T47528  
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24468  
A:Accession: T47528  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-431 <TOR>  
A:Cross-references: EMBL:AL162459  
A:Experimental source: cultivar Columbia; BAC clone F16L2  
C:Genetics:  
A:Map position: 3  
A:Introns: 216/3: 275/2: 299/1: 342/3  
A:Note: F16L2.90  
C:Superfamily: Drosophila hypothetical protein CG10133

Query Match 77.6%; Score 45; DB 2; Length 431;  
Best Local Similarity 85.7%; Pred. No. 2.7;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GPRPFHC 7  
DB 259 GPRPFHC 265

## RESULT 3

S52390  
D:hordein precursor - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 08-May-1995 #sequence\_revision 26-Jul-1996 #text\_change 20-Aug-1999  
C:Accession: S52390  
R:Sorensen, M.B.; Muller, M.; Simpson, D.  
submitted to the EMBL Data Library, February 1995  
A:Description: Hordein promoter methylation and transcriptional activity in wild type and  
A:Reference number: S52390  
A:Accession: S52390  
A:Molecule type: DNA  
A:Residues: 1-475 <SOR>  
A:Cross-references: EMBL:X84368; NID:g671536; PIDN:CAA59104.1; PID:g671537  
R:Hallford, N.G.; Tatham, A.S.; Sul, E.; Daroda, L.; Dreyer, T.; Shewry, P.R.  
Biochem. Biophys. Acta 1122, 118-122, 1992  
A:Title: Identification of a novel beta-turn-rich repeat motif in the D hordeins of barl  
A:Reference number: S23921; MUID:92353095  
A:Accession: S23921  
A:Molecule type: mRNA  
A:Residues: 267-355, 'P', 357-359, 'Q', 361-458, 'Y', 460-707 <HAD>  
A:Cross-references: EMBL:X68072; NID:g18969; PIDN:CAA48209.1; PID:g18970  
C:Genetics:  
A:Gene: hor3  
C:Superfamily: glutenin  
C:Keywords: seed; storage protein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-707/Product: D-hordein #status predicted <MAT>

Query Match 72.4%; Score 42; DB 2; Length 707;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
DB 1 GPRPFHCQ 9

DB 528 GKPFHCQ 536

## RESULT 4

TVCHLV  
epidermal growth factor receptor precursor - chicken  
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
C:Species: Gallus gallus (chicken)  
C:Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000  
C:Accession: A27720; A00643  
R:Iax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Venn  
Mol. Cell Biol. 8, 1970-1978, 1988  
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in  
A:Reference number: A27720; MUID:88261272  
A:Accession: A27720  
A:Molecule type: mRNA  
A:Residues: 1-1223 <LAX>  
A:Cross-references: GB:M20386  
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines  
Cell 41, 719-726, 1985  
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and  
A:Reference number: A00643; MUID:85228222  
A:Accession: A00643  
A:Molecule type: mRNA  
A:Residues: 585-1223 <NLT>  
A:Cross-references: GB:M10066  
C:Genetics:  
A:Gene: erbB  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth fact  
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth fact  
specific protein kinase  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
F:31-654/Domain: extracellular #status predicted <EXT>  
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>  
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>  
F:655-677/Domain: transmembrane #status predicted <TM>  
F:678-1223/Domain: intracellular #status predicted <INT>  
F:719-984/Domain: protein kinase homology <KIN>  
F:727-735/Region: protein kinase ATP-binding motif  
F:136-202, 280, 361, 370, 422, 575, 580, 615, 635/Binding site: carbohydrate (Thr) (covalent)  
F:162, 650/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:667/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicte  
F:754/Active site: Lys #status predicted  
F:1100, 1183, 1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #s

Query Match 70.7%; Score 41; DB 1; Length 1223;  
Best Local Similarity 85.7%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPFHC 7  
DB 524 GPRPFHC 530

## RESULT 5

S44281  
ES43 protein - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Mar-1999  
R:Speulman, E.; Salamin, F.  
submitted to the EMBL Data Library, February 1994  
A:Reference number: S44281  
A:Accession: S44281  
A:Molecule type: mRNA  
A:Residues: 1-227 <SPE>  
A:Cross-references: EMBL:X77575; NID:g482918; PID:g13445528  
C:Genetics:  
A:Gene: ES43

Query Match 69.0%; Score 40; DB 2; Length 227;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
1 1 1 1 1 1  
DB 79 GRRPFHCER 87

## RESULT 6

S72766  
hypothetical protein B1496\_C3\_217 - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001  
C:Accession: S72766; T11018  
R:Smith, D.R.; Robison, K.  
Submitted to the EMBL Data Library, November 1993  
A:Description: Mycobacterium leprae cosmid B1496.  
A:Reference number: S72695  
A:Accession: S72766  
A:Molecule type: DNA  
A:Residues: 1-101 <SMT>  
A:Cross-References: EMBL:U00013; NID:g466868; PIDN:AAAI7133.1; PID:g466888  
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, September 1997  
A:Reference number: Z16918  
A:Accession: T11018  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-101 <PAR>  
A:Cross-References: EMBL:Z99125; NID:g2398683; PIDN:CABI6179.1; PID:g343553; PID:g239871  
C:Genetics:  
A:Gene: MLCU536\_35c  
A:Start codon: GTG

Query Match 67.2%; Score 39; DB 2; Length 101;  
Best Local Similarity 66.7%; Pred. No. 7.9;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
1 1 1 1 1 1  
DB 80 GPTPFSCRR 88

## RESULT 7

T08009  
probable ribosomal protein L5 - green alga (Dunaliella salina)  
C:Species: Dunaliella salina  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999  
C:Accession: T08009  
R:KO, J.H.; Lee, S.H.  
Submitted to the EMBL Data Library, October 1997  
A:Description: A cDNA encodes a protein sequence homologous to the eukaryotic ribosomal  
A:Reference number: Z16285  
A:Accession: T08009  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-271 <KOU>  
A:Cross-References: EMBL:AF028833; NID:g2599103; PIDN:AAB84056.1; PID:g2599104  
C:Genetics:  
A:Gene: RPL  
C:Superfamily: rat ribosomal protein L5  
C:Keywords: ribosome

Query Match 67.2%; Score 39; DB 2; Length 271;  
Best Local Similarity 71.4%; Pred. No. 19;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GPRPFHC 7  
1 1 1 1 1 1  
DB 138 GPKPFYC 144

RESULT 8  
S33701  
homeotic protein Dch-1 - planarian (Dugesia tigrina)

C:Species: Dugesia tigrina  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Sep-1999  
C:Accession: S33701; A41151  
R:Garcia-Fernandez, J.; Baguna, J.; Salo, E.  
Development 118, 241-253, 1993  
A:Title: Genomic organization and expression of the planarian homeobox genes Dch-1 an  
A:Reference number: S33701; MUID:93387216  
A:Accession: S33701  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-533 <GAR>  
A:Cross-References: EMBL:X69200  
R:Garcia-Fernandez, J.; Baguna, J.; Salo, E.  
Proc. Natl. Acad. Sci. U.S.A. 88, 7338-7342, 1991  
A:Title: Planarian homeobox genes: cloning, sequence analysis, and expression.  
A:Reference number: A41151; MUID:91334461  
A:Accession: A41151  
A:Molecule type: mRNA  
A:Residues: 83-533 <GA2>  
A:Cross-References: GB:X56499; NID:g9155; PIDN:CAA39854.1; PID:g9156  
A:Experimental source: intact and regenerating planarians  
C:Genetics:  
A:Introns: 135/3; 271/1  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:378-434/Domain: homeobox homology <HOX>

Query Match 67.2%; Score 39; DB 2; Length 533;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
1 1 1 1 1 1  
DB 228 GPRLLHCQ 236

## RESULT 9

B34612  
zinc finger protein ZNF8 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 01-Dec-2000  
C:Accession: B34612  
R:Ianata, L.; Dotti, E.; Pannuti, A.; Pascucci, A.; Penque, G.; Feliciello, I.; La Man  
Genomics 6, 333-340, 1990  
A:Title: cDNA isolation, expression analysis, and chromosomal localization of two hum  
A:Reference number: A34612; MUID:90169993  
A:Accession: B34612  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-543 <LAN>  
A:Cross-References: GB:M2581; NID:g340447; PIDN:AAAG1314.1; PID:g340448; GB:J04751  
C:Genetics:  
A:Gene: GDB:ZNF8  
A:Cross-References: GDB:120510; OMIM:194532  
A:Map position: 20q13-20q13  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C:Keywords: DNA binding; zinc finger

Query Match 67.2%; Score 39; DB 2; Length 543;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCQ 9  
1 1 1 1 1 1  
DB 389 GEPFECRQ 397

RESULT 10  
KIR7C  
protein kinase C (EC 2.7.1.-) gamma - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1992 #sequence,revision 30-Sep-1992 #text\_change 11-Jun-1999  
C:Accession: A05105; S02129; I55317  
R:Knopf, J.L.; Lee, M.H.; Sultzman, L.A.; Kriz, R.W.; Loomis, C.R.; Hewick, R.M.; Bell, Cell 46, 491-502, 1986  
A:Title: Cloning and expression of multiple protein kinase C cDNAs.  
A:Reference number: A90883; MUID:86272097  
A:Accession: A05105  
A:Molecule type: mRNA  
A:Residues: 1-697 <KNO>  
A:Cross-references: GB:M13707; NID:g206186; PIDN:AAA41874.1; PID:g206187  
A:Note: the authors translated the codon UUU for residue 432 as Glu  
R:Ono, Y.; Fujii, T.; Igarashi, K.; Kikkawa, U.; Ogita, K.; Nishizuka, Y. Nucleic Acids Res. 16, 5199-5200, 1988  
A:Title: Nucleotide sequences of cDNAs for alpha and gamma subspecies of rat brain prote  
A:Reference number: S02129; MUID:88262515  
A:Accession: S02129  
A:Molecule type: mRNA  
A:Residues: 1-697 <ONO>  
A:Cross-references: EMBL:X07287; NID:g56917; PIDN:CAA30267.1; PID:g56918  
R:Chen, K. J. Biol. Chem. 265, 19961-19965, 1990  
A:Title: Characterization of the 5'-flanking region of the rat protein kinase C gamma ge  
A:Reference number: I55317; MUID:91060619  
A:Accession: I55317  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-56 <RES>  
A:Cross-references: GB:M55417; NID:g206184; PIDN:AAA41873.1; PID:g554487  
C:Comment: This is a calcium-activated, phospholipid-dependent, serine- and threonine-S  
I inositol phospholipids. This protein is a receptor for tumor-promoting phorbol esters.  
C:Comment: Binding to acidic phospholipids (phosphatidylserine) in the cell membrane may  
C:Comment: The zinc-stabilized regions bind diacylglycerol and phorbol esters.  
C:genetics:  
A:gene: PKC-gamma  
C:Superfamily: protein kinase C alpha; protein kinase C C2 region homology; protein kin  
C:Keywords: ATP; autophosphorylation; calcium binding; duplication; phorbol ester bindin  
F:18-28/Region: pseudophosphorylation motif  
F:21-26/Region: pseudophosphorylation motif  
F:36-85/Domain: protein kinase C zinc-binding repeat homology <KZ1>  
F:101-150/Domain: protein kinase C zinc-binding repeat homology <KZ2>  
F:151-264/Domain: protein kinase C C2 region homology <KC2>  
F:349-614/Domain: protein kinase homology <KIN>  
F:357-365/Region: protein kinase ATP-binding motif  
F:36,66,69,85/Binding site: zinc (His, Cys, Cys, Cys) #status predicted  
F:49,52,74,77/Binding site: zinc (His, Cys, His, Cys) #status predicted  
F:101,131,134,150/Binding site: zinc (His, Cys, Cys, Cys) #status predicted  
F:114,117,139,142/Binding site: zinc (His, Cys, His, Cys) #status predicted  
F:380/Active site: Lys #status predicted  
F:648,655/Binding site: phosphate (Thr) (covalent) (by autophosphorylation) #status pred

Query Match 67.2%; Score 39; DB 1; Length 697;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRO 9  
IIII II:  
Db 14 GPRPLFCRK 22

RESULT 11  
KIRBGC  
protein kinase C (EC 2.7.1.-) gamma - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 30-Sep-1992 #sequence,revision 30-Sep-1992 #text\_change 11-Jun-1999  
C:Accession: A28708  
R:Ohno, S.; Kawasaki, H.; Kono, Y.; Inagaki, M.; Hidaka, H.; Suzuki, K. Biochemistry 27, 2083-2087, 1988

A:Title: A fourth type of rabbit protein kinase C.  
A:Reference number: A28708; MUID:88241036  
A:Accession: A28708  
A:Molecule type: mRNA  
A:Residues: 1-697 <KNO>  
A:Cross-references: GB:M19338; NID:g165651; PIDN:AAA31449.1; PID:g165652  
C:Comment: This is a calcium-activated, phospholipid-dependent, serine- and threonine  
I inositol phospholipids. This protein is a receptor for tumor-promoting phorbol este  
C:Comment: Binding to acidic phospholipids (phosphatidylserine) in the cell membrane  
C:Comment: The zinc-stabilized regions bind diacylglycerol and phorbol esters.  
C:Superfamily: protein kinase C alpha; protein kinase C C2 region homology; protein k  
C:Keywords: ATP; autophosphorylation; calcium binding; duplication; phorbol ester bin  
F:18-28/Region: pseudophosphorylation motif  
F:21-26/Region: pseudophosphorylation motif  
F:36-85/Domain: protein kinase C zinc-binding repeat homology <KZ1>  
F:101-150/Domain: protein kinase C zinc-binding repeat homology <KZ2>  
F:151-264/Domain: protein kinase C C2 region homology <KC2>  
F:349-614/Domain: protein kinase homology <KIN>  
F:357-365/Region: protein kinase ATP-binding motif  
F:36,66,69,85/Binding site: zinc (His, Cys, Cys, Cys) #status predicted  
F:49,52,74,77/Binding site: zinc (His, Cys, His, Cys) #status predicted  
F:101,131,134,150/Binding site: zinc (His, Cys, Cys, Cys) #status predicted  
F:114,117,139,142/Binding site: zinc (His, Cys, His, Cys) #status predicted  
F:380/Active site: Lys #status predicted  
F:648,655/Binding site: phosphate (Thr) (covalent) (by autophosphorylation) #status p

Query Match 67.2%; Score 39; DB 1; Length 697;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRO 9  
IIII II:  
Db 14 GPRPLFCRK 22

RESULT 12  
JN0548  
protein kinase C (EC 2.7.1.-) gamma - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence,revision 31-Dec-1993 #text\_change 18-Jun-1999  
C:Accession: JN0548; S20735; S29620  
R:Bowers, B.L.; Parham, C.L.; Sikelja, J.M.; Wehner, J.M. Gene 123, 263-265, 1993  
A:Title: Isolation and sequence of a mouse brain cDNA coding for protein kinase C-gam  
A:Reference number: JN0548; MUID:93154595  
A:Accession: JN0548  
A:Molecule type: mRNA  
A:Residues: 1-697 <BOW>  
A:Cross-references: EMBL:X67129; NID:g53696; PIDN:CAA47608.1; PID:g53697  
A:Experimental source: brain  
R:Leitges, M.; Proikas, T.; Strabel, S. submitted to the EMBL Data Library, April 1992  
A:Description: Comparison of the mouse and rat protein kinase C gamma gene promoter.  
A:Reference number: S20735  
A:Accession: S20735  
A:Molecule type: DNA  
A:Residues: 1-67 <LEI>  
A:Cross-references: EMBL:X65720; NID:g53694; PIDN:CAA46636.1; PID:g53695  
C:genetics:  
A:introns: 57/2  
C:Superfamily: protein kinase C alpha; protein kinase C C2 region homology; protein k  
C:Keywords: ATP; autophosphorylation; calcium binding; duplication; phorbol ester bin  
F:36-85/Domain: protein kinase C zinc-binding repeat homology <KZ1>  
F:101-150/Domain: protein kinase C zinc-binding repeat homology <KZ2>  
F:151-264/Domain: protein kinase C C2 region homology <KC2>  
F:349-614/Domain: protein kinase homology <KIN>  
F:357-365/Region: protein kinase ATP-binding motif

Query Match 67.2%; Score 39; DB 2; Length 697;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
 1111 11:  
 Db 14 GPRPLFCRK 22

RESULT 13  
 D24664  
 protein kinase C (EC 2.7.1.-) gamma - human  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1988 #sequence\_revision 10-May-1996 #text\_change 23-Mar-2001  
 C:Accession: D24664; S31611; S36837  
 R:Conseus, L.; Parker, P.J.; Rhee, L.; Yang-Peng, T.L.; Chen, E.; Waterfield, M.D.; Fra  
 Science 233, 859-866, 1986  
 A:Title: Multiple, distinct forms of bovine and human protein kinase C suggest diversity  
 A:Reference number: A94291; MUID:86289426  
 A:Accession: D24664  
 A:Molecule type: mRNA  
 A:Residues: 1-313, 'VS', '316', 'T', '318 <COU>  
 A>Note: the authors translated the codon GCA for residue 170 as Ser and AGC for residue  
 R: Hug, H.  
 submitted to the EMBL Data Library, September 1992  
 A:Description: Partial cDNA sequence of human protein kinase C zeta.  
 A:Reference number: S25605  
 A:Accession: S31611  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 162-697 <HUG>  
 A:Cross-references: EMBL:215114; NID:g35496; PIDN:CAA78820.1; PID:g35497  
 R:Kochs, G.; Hummel, R.; Meyer, D.; Hug, H.; Marne, D.; Sarre, T.F.  
 Eur. J. Biochem. 216, 597-606, 1993  
 A:Title: Activation and substrate specificity of the human protein kinase C alpha and zeta  
 A:Reference number: S36836; MUID:93387312  
 A:Accession: S36837  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 162-697 <KOC>  
 A:Cross-references: EMBL:215114; NID:g35496; PIDN:CAA78820.1; PID:g35497  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992  
 C:Genetics:  
 A:Gene: GDB:PRKCG  
 A:Cross-references: GDB:128017; OMIM:176980  
 A:Map position: 19q13.4-19q13.4  
 C:Superfamily: protein kinase C alpha: protein kinase C C2 region homology; protein kin  
 C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; serine/threon  
 F:36-85/Domain: protein kinase C zinc-binding repeat homology <KZ1>  
 F:101-150/Domain: protein kinase C zinc-binding repeat homology <KZ2>  
 F:151-264/Domain: protein kinase C C2 region homology <KIN>  
 F:349-614/Domain: protein kinase homology <KIN>  
 F:357-365/Region: protein kinase ATP-binding motif

Query Match 67.2%; Score 39; DB 2; Length 697;  
 Best Local Similarity 66.7%; Pred. No. 44;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
 1111 11:  
 Db 14 GPRPLFCRK 22

RESULT 14  
 B57785  
 zinc finger protein ZNF136 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 01-Dec-2000  
 C:Accession: B57785; S66507; S66510  
 R:Tommerup, N.; Vissing, H.  
 Genomics 27, 259-264, 1995  
 A:Title: Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identifi  
 A:Reference number: A57785; MUID:96044430  
 A:Accession: B57785  
 A>Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-540 <TOM>  
 A:Cross-references: GB:U09367; NID:9487784; PIDN:AA050261.1; PID:9487785  
 R:Vissing, H.; Meyer, W.K.H.; Aagaard, L.; Tommerup, N.; Thiesen, H.J.  
 FEBS Lett. 369, 153-157, 1995  
 A:Title: Repression of transcriptional activity by heterologous KRAB domains present  
 A:Reference number: S66506; MUID:95377390  
 A:Accession: S66507  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-82 <VTS>  
 A:Cross-references: EMBL:U09367  
 C:Genetics:  
 A:Gene: GDB:ZNF136  
 A:Cross-references: GDB:137035  
 A:Map position: 19p13.2-19p13.12  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 65.5%; Score 38; DB 2; Length 540;  
 Best Local Similarity 44.8%; Pred. No. 52;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 9  
 1111 11:  
 Db 500 GOKPFHCKE 508

RESULT 15  
 T13594  
 hypothetical protein peb - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
 C:Accession: T13594  
 R:Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
 A:Reference number: Z17692  
 A:Accession: T13594  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1891 <FER>  
 A:Cross-references: EMBL:AL031227; NID:e1330103; PID:e1316856; PIDN:CAA20227.1  
 C:Genetics:  
 A:Gene: peb  
 A:Cross-references: FlyBase:FBgn0003053  
 A:Introns: 289/3

Query Match 65.5%; Score 38; DB 2; Length 1891;  
 Best Local Similarity 75.0%; Pred. No. 1,6e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPRPFHCRO 8  
 1111 11:  
 Db 306 GERPFMCR 313

Search completed: May 24, 2002, 16:50:12  
 Job time: 317 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:59 ; Search time 49.37 Seconds

(without alignments)  
4.453 Million cell updates/sec

Title: US-09-730-379E-3

Perfect score: 58

Sequence: 1 GPRFHCRQ 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCROS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	67.2	334	2	US-08-665-647-3
2	39	67.2	1050	3	US-09-045-632-49
3	39	67.2	1050	3	US-09-045-632-50
4	37	63.8	56	4	US-08-711-417C-186
5	37	63.8	56	4	US-08-711-417C-187
6	37	63.8	56	4	US-08-711-417C-188
7	37	63.8	56	4	US-08-711-417C-189
8	37	63.8	334	4	US-08-711-417C-201
9	37	63.8	334	4	US-08-711-417C-199
10	37	63.8	431	4	US-08-711-417C-195
11	37	63.8	432	4	US-08-711-417C-197
12	37	63.8	461	4	US-08-711-417C-196
13	37	63.8	470	2	US-08-465-590-153
14	37	63.8	470	4	US-08-283-300A-13
15	37	63.8	470	4	US-08-711-417C-153
16	37	63.8	470	5	PCT-US95-09345-13
17	37	63.8	516	4	US-08-711-417C-202
18	37	63.8	518	4	US-08-711-417C-198
19	37	63.8	568	5	PCT-US93-08743-5
20	37	63.8	708	1	US-08-396-479B-8
21	37	63.8	708	1	US-08-818-823-8
22	37	63.8	739	1	US-08-396-479B-10
23	37	63.8	739	1	US-08-818-823-10
24	37	63.8	1068	1	US-08-396-479B-12
25	37	63.8	1068	1	US-08-818-823-12
26	37	63.8	1075	5	PCT-US94-07297-41
27	36	62.1	410	3	US-09-083-521-2

28	35	60.3	69	2	US-08-726-306A-53	Sequence 53, Appl
29	35	60.3	127	4	US-09-199-637A-275	Sequence 275, App
30	35	60.3	338	2	US-08-933-750C-4	Sequence 4, Appl1
31	35	60.3	338	4	US-09-234-613-4	Sequence 4, Appl1
32	35	60.3	706	1	US-08-074-967-2	Sequence 2, Appl1
33	35	60.3	706	2	US-08-553-541B-2	Sequence 2, Appl1
34	35	60.3	706	5	US-09-268-202-2	Sequence 2, Appl1
35	35	60.3	706	5	PCT-US94-06669-2	Sequence 2, Appl1
36	35	60.3	803	4	US-09-063-035-2	Sequence 2, Appl1
37	34	58.6	10	1	US-08-040-548-26	Sequence 26, Appl
38	34	58.6	10	1	US-08-466-344-26	Sequence 26, Appl
39	34	58.6	28	4	US-09-253-396A-229	Sequence 229, Appl
40	34	58.6	28	4	US-09-037-179B-15	Sequence 15, Appl
41	34	58.6	59	1	US-08-040-548-7	Sequence 7, Appl1
42	34	58.6	59	1	US-08-466-344-7	Sequence 7, Appl1
43	34	58.6	86	6	5206152-3	Patent No. 5206152
44	34	58.6	89	1	US-08-040-548-8	Sequence 8, Appl1
45	34	58.6	89	1	US-08-466-344-8	Sequence 8, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-665-647-3  
Sequence 3, Application US/08665647  
Patent No. 5935803  
GENERAL INFORMATION:  
APPLICANT: Dasquez, Nicki J.  
APPLICANT: Ron, Dorit  
APPLICANT: Voronova, Anna F.  
APPLICANT: Napolitano, Eugene W.  
TITLE OF INVENTION: METHODS TO IDENTIFY IMMUNOMODULATORS  
TITLE OF INVENTION: USING COGNATE INTERACTION OF PKC-THETA  
NUMBER OF SEQUENCES: 89  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue, NW - Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,647  
FILING DATE: 18-JUN-1996  
CLASSIFICATION: A35  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22550-20025.25  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 822-0168  
TELEX: 90-4030 MRSNFOERSWSH  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 334 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-665-647-3

Query Match 67.2%; Score 39; DB 2; Length 334;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 PRPFHCRQ 9

Db 172 PRPFCNO 179

## RESULT 2

US-09-045-632-49  
Sequence 49, Application US/09045632  
Patent No. 6001575  
GENERAL INFORMATION:  
APPLICANT: Huganir, Richard L.  
APPLICANT: Dong, Hualing  
TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
City: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/045,632  
FILING DATE: 19-MAR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,016  
FILING DATE: 19-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Corless, Peter F.  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 48147/1699-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1050 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-045-632-49

Query Match 67.2%; Score 39; DB 3; Length 1050;  
Best Local Similarity 85.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 PRPFCR 8  
Db 405 PRPFCR 411

RESULT 3  
US-09-045-632-50  
Sequence 50, Application US/09045632  
Patent No. 6001575  
GENERAL INFORMATION:  
APPLICANT: Huganir, Richard L.  
APPLICANT: Dong, Hualing  
TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
City: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/045,632  
FILING DATE: 19-MAR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,016  
FILING DATE: 19-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Corless, Peter F.  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 48147/1699-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1050 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-045-632-50

Query Match 67.2%; Score 39; DB 3; Length 1050;  
Best Local Similarity 85.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 PRPFCR 8  
Db 405 PRPFCR 411

RESULT 4  
US-08-711-417C-186  
Sequence 186, Application US/08711417C  
Patent No. 6228611  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF SEQUENCES: 202  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
City: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-Sep-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965



REFERENCE/DOCKET NUMBER: 10287/007001  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 186:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 186:

US-08-711-417C-186

Query Match 63.8%; Score 37; DB 4; Length 56;

Best Local Similarity 66.7%; Pred. No. 9.8;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9

Db 1 GERPFQCNQ 9

RESULT 5

US-08-711-417C-187

Sequence 187, Application US/08711417C

Patent No. 6228611

GENERAL INFORMATION:

APPLICANT: Georgopoulos, Katia A.

TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE

NUMBER OF SEQUENCES: 202

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA: US/08/711,417C

APPLICATION NUMBER: 08/238,212

FILING DATE: 05-Sep-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/238,212

FILING DATE: 02-May-1994

APPLICATION NUMBER: 08/121,438

FILING DATE: 14-Sep-1993

APPLICATION NUMBER: 07/946,233

FILING DATE: 14-Sep-1992

ATTORNEY/AGENT INFORMATION:

NAME: Myers, Louis P.

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: 10287/007001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 187:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 187:

US-08-711-417C-187

Query Match 63.8%; Score 37; DB 4; Length 56;

Best Local Similarity 66.7%; Pred. No. 9.8;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9

Db 1 GERPFQCNQ 9

RESULT 6

US-08-711-417C-188

Sequence 188, Application US/08711417C

Patent No. 6228611

GENERAL INFORMATION:

APPLICANT: Georgopoulos, Katia A.

TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE

NUMBER OF SEQUENCES: 202

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/711,417C

FILING DATE: 05-Sep-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/238,212

FILING DATE: 02-May-1994

APPLICATION NUMBER: 08/121,438

FILING DATE: 14-Sep-1993

APPLICATION NUMBER: 07/946,233

FILING DATE: 14-Sep-1992

ATTORNEY/AGENT INFORMATION:

NAME: Myers, Louis P.

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: 10287/007001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 188:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 188:

US-08-711-417C-188

Query Match 63.8%; Score 37; DB 4; Length 56;

Best Local Similarity 66.7%; Pred. No. 9.8;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9

Db 1 GERPFQCNQ 9

RESULT 7

US-08-711-417C-189

Sequence 189, Application US/08711417C

Patent No. 6228611

GENERAL INFORMATION:

APPLICANT: Georgopoulos, Katia A.

TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE

NUMBER OF SEQUENCES: 202

SEQUENCE DESCRIPTION: SEQ ID NO: 189:

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-SEP-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10287/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 189:  
US-08-711-417C-189

Query Match 63.8%; Score 37; DB 4; Length 56;  
Best Local Similarity 66.7%; Pred. No. 9.8;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

Qy 1 GPRPHCRQ 9  
| | | | |  
Db 1 GERPQCNO 9

RESULT 8  
US-08-711-417C-201  
Sequence 201 Application US/08711417C  
Patent No. 6228611  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF SEQUENCES: 202  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-SEP-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10287/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 189:  
US-08-711-417C-189

```

1      APPLICATION NUMBER: 08/238,212
2      FILING DATE: 02-MAY-1994
3      APPLICATION NUMBER: 08/121,438
4      FILING DATE: 14-SEP-1993
5      APPLICATION NUMBER: 07/946,233
6      FILING DATE: 14-SEP-1992
7      ATTORNEY/AGENT INFORMATION:
8      NAME: Myers, Louis P.
9      REGISTRATION NUMBER: 35,965
10     REFERENCE/DOCKET NUMBER: 10287/007001
11     TELECOMMUNICATION INFORMATION:
12     TELEPHONE: 617/542-5070
13     TELEFAX: 617/542-8906
14     TELE: 200154
15     SEQUENCE CHARACTERISTICS:
16     LENGTH: 334 amino acids
17     TYPE: amino acid
18     TOPOLOGY: linear
19     MOLECULE TYPE: protein
20     FRAGMENT TYPE: internal
21     SEQUENCE DESCRIPTION: SEQ ID NO: 201:
22     US-08-711-417C-201
23
24     Query Match          63.8%; Score 37; DB 4; Length 334;
25     Best Local Similarity 66.7%; Pred. No. 51;
26     Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
27
28     QY      1 GPRPFCRO 9
29             | | | | |
30     Db      1 GERPFCNO 9
31
32 RESULT      9
33 US-08-711-417C-199
34 ; Sequence 199, Application US/08711417C
35 ; Patent No. 6228611
36 ;
37 ; GENERAL INFORMATION:
38 ;
39 ; APPLICANT: Georgopoulos, Katia A.
40 ; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
41 ; NUMBER OF SEQUENCES: 202
42 ; CORRESPONDENCE ADDRESS:
43 ; ADDRESSEE: Fish & Richardson P.C.
44 ; STREET: 225 Franklin Street
45 ; CITY: Boston
46 ; STATE: MA
47 ; COUNTRY: USA
48 ; ZIP: 02110-2804
49 ;
50 ; COMPUTER READABLE FORM:
51 ; MEDIUM TYPE: Diskette
52 ;
53 ; COMPUTER: IBM Compatible
54 ; OPERATING SYSTEM: Windows 95
55 ; SOFTWARE: FASTSEQ for Windows Version 2.0b
56 ; CURRENT APPLICATION DATA:
57 ; APPLICATION NUMBER: US/08/711,417C
58 ; FILING DATE: 05-Sep-1996
59 ; PRIOR APPLICATION DATA:
60 ; APPLICATION NUMBER: 08/238,212
61 ; FILING DATE: 02-MAY-1994
62 ; APPLICATION NUMBER: 08/121,438
63 ; FILING DATE: 14-SEP-1993
64 ; APPLICATION NUMBER: 07/946,233
65 ; FILING DATE: 14-SEP-1992
66 ; ATTORNEY/AGENT INFORMATION:
67 ;
68 ; NAME: Myers, Louis P.
69 ; REGISTRATION NUMBER: 35,965
70 ; REFERENCE/DOCKET NUMBER: 10287/007001
71 ; TELECOMMUNICATION INFORMATION:
72 ; TELEPHONE: 617/542-5070
73 ; TELEFAX: 617/542-8906
74 ; TELE: 200154
75 ;
76 ; INFORMATION FOR SEQ ID NO: 199:
77 ;
78 ; TELEPHONE: 617/542-5070
79 ; TELEFAX: 617/542-8906
80 ; TELE: 200154
81 ;
82 ; INFORMATION FOR SEQ ID NO: 199:

```

SEQUENCE CHARACTERISTICS:  
LENGTH: 390 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 199;  
US-08-711-417C-199

Query Match 63.8%; Score 37; DB 4; Length 390;  
Best Local Similarity 66.7%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9  
| | | | |  
DB 54 GERPFQCNQ 62

RESULT 10  
US-08-711-417C-195  
; Sequence 195, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/711,417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 08/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 195:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 195;  
US-08-711-417C-195

Query Match 63.8%; Score 37; DB 4; Length 431;  
Best Local Similarity 66.7%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GPRPFHCQ 9

DB 54 GERPFQCNQ 62  
| | | | |

RESULT 11  
US-08-711-417C-197  
; Sequence 197, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/711,417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 08/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 197:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 432 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 197;  
US-08-711-417C-197

Query Match 63.8%; Score 37; DB 4; Length 432;  
Best Local Similarity 66.7%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9  
| | | | |  
DB 141 GERPFQCNQ 149

RESULT 12  
US-08-711-417C-196  
; Sequence 196, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street

CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-Sep-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10287/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 196:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 461 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 196:  
US-08-711-417C-196

Query Match 63.8%; Score 37; DB 4; Length 461;  
Best Local Similarity 66.7%; Pred. No. 69;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFCRQ 9  
DB 86 GERPFQCNQ 94

RESULT 13  
US-08-465-590-153  
Sequence 153, Application US/08465590  
Patent No. 5824770  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF SEQUENCES: 164  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, Suite 510  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,590  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/238,212  
FILING DATE: 02-MAY-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/121,438  
FILING DATE: 14-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,695  
REFERENCE/DOCKET NUMBER: MPG-006C2DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 153:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: C-terminal  
US-08-465-590-153

Query Match 63.8%; Score 37; DB 2; Length 470;  
Best Local Similarity 66.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPRPFCRQ 9  
DB 90 GERPFQCNQ 98

RESULT 14  
US-08-283-300A-13  
Sequence 13, Application US/08283300A  
Patent No. 6172278  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS TRANSGENIC CELLS AND ANIMALS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, Suite 510  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,300A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/238,212  
FILING DATE: 02-MAY-94  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/121,438  
FILING DATE: 14-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,695  
REFERENCE/DOCKET NUMBER: MGP-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-5941  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: C-terminal  
US-08-283-300A-13

Query Match 63.8%; Score 37; DB 4; Length 470;  
Best Local Similarity 66.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GPRPFHCQ 9  
Db 90 GERPFQCQ 98

RESULT 15

US-08-711-417C-153  
; Sequence 153, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/711,417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 08/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 153  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 470 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: C-terminal  
; FEATURE:  
; NAME/KEY: Other  
; LOCATION: 1...470  
; OTHER INFORMATION: Xaa = any amino acid  
; SEQUENCE DESCRIPTION: SEQ ID NO: 153:  
US-08-711-417C-153

Query Match 63.8%; Score 37; DB 4; Length 470;

Best Local Similarity 66.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GPRPFHCQ 9  
Db 90 GERPFQCQ 98

Search completed: May 24, 2002, 16:49:00  
Job time: 285 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 16:48:03 ; Search time 136.41 Seconds  
(without alignments)  
7.328 Million cell updates/sec

Title: US-09-730-379E-3  
Perfect score: 58  
Sequence: 1 GPRPFHCRCQ 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	77.6	386	22 AAM93303	Human polypeptide,
2	45	77.6	500	20 AAY34147	Murine Helios-2 pr
3	45	77.6	526	20 AAY34146	Murine Helios-1 pr
4	45	77.6	526	20 AAY34148	Murine Helios-2 pr
5	45	77.6	533	21 AAY87745	Murine embryonic b
6	45	77.6	725	22 AAB95116	Human protein sequ
7	43	74.1	19	20 AAY01161	Polypeptide fragme
8	43	74.1	259	20 AAY01159	Polypeptide fragme
9	43	74.1	259	21 AAY73354	HTRM clone 191094
10	42	72.4	121	22 AAU15915	Human novel secret
11	42	72.4	529	22 AAU28084	Novel human secret

12	41	70.7	162	22 AAU61505	Propionibacterium
13	40	69.0	53	22 AAU50534	Propionibacterium
14	40	69.0	76	21 AAG03279	Human secreted pro
15	40	69.0	140	22 AAU16426	Human novel secret
16	40	69.0	141	22 AAU32943	Novel human secret
17	40	69.0	369	22 ABO1477	Novel human diagno
18	40	69.0	498	22 AAB94669	Human protein sequ
19	40	69.0	566	22 ABO8714	Novel human diagno
20	40	69.0	619	22 AAB95103	Human protein sequ
21	40	69.0	714	22 AAU32283	Novel human secret
22	39	67.2	69	22 AAM85477	Human immune/haema
23	39	67.2	141	22 ABB11039	Human Zn finger pr
24	39	67.2	192	22 AAU16093	Human novel secret
25	39	67.2	192	22 AAU16520	Human novel secret
26	39	67.2	280	21 AAB19565	Rat mutated protei
27	39	67.2	281	22 AAO12907	Human polypeptide
28	39	67.2	335	18 AAW15771	Protein cognate of
29	39	67.2	383	22 ABO10467	Novel human diagno
30	39	67.2	400	22 ABB59616	Drosophila melanog
31	39	67.2	575	22 AAU27687	Human full-length
32	39	67.2	696	21 AAB19564	Rat protein kinase
33	39	67.2	697	9 AAP82018	Protein kinase C-I-
34	39	67.2	697	22 AAM78746	Human protein SEQ
35	39	67.2	730	22 AAM79730	Human protein SEQ
36	39	67.2	760	22 ABO22744	Novel human diagno
37	39	67.2	1049	19 AAM73062	Rat GRIP2. Rattus
38	39	67.2	1196	22 ABO28743	Novel human diagno
39	38	65.5	60	22 AAU49374	Propionibacterium
40	38	65.5	80	22 AAU16436	Human novel secret
41	38	65.5	509	22 AAM93200	Human polypeptide,
42	38	65.5	1893	22 ABB59829	Drosophila melanog
43	37	63.8	25	20 AAY26277	Isolated peptide f
44	37	63.8	25	20 AAY26278	Isolated peptide f
45	37	63.8	25	20 AAY26279	Isolated peptide f

# ALIGNMENTS

RESULT 1	
AAM93303	
ID AAM93303 standard; Protein; 386 AA.	
XX AC AAM93303;	
XX DT 06-NOV-2001 (first entry)	
XX DE Human polypeptide, SEQ ID NO: 2805.	
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.	
XX OS Homo sapiens.	
XX PN EP130094-A2.	
XX PD 05-SEP-2001.	
XX PF 07-JUL-2000; 2000EP-0114089.	
XX PR 08-JUL-1999; 99JP-0194486.	
XX PR 11-JAN-2000; 2000JP-0118774.	
XX PR 02-MAY-2000; 2000JP-0183765.	
XX PA (HELI-) HELIX RES INST.	
XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX DR WPI: 2001-524255/58.	
XX DR N-PSDB; AAK94223.	
XX PT 830 Primers useful for synthesizing full length cDNA clones and their	
PT use in genetic manipulation -	

XX Claim 8; SEQ ID NO 2805; 1380pp + sequence listing; English.  
 XX The invention relates to primers for synthesising full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesising the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a polypeptide  
 CC encoded by a full length human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.  
 XX Sequence 386 AA;  
 SQ

Query Match 77.6%; Score 45; DB 22; Length 386;  
 Best Local Similarity 77.8%; Pred. No. 6.7;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRCQ 9  
 Db 208 gprphhctq 216  
 ||||| | |

RESULT 2  
 AAY34147  
 ID AAY34147 standard; Protein; 500 AA.  
 AC AAY34147;  
 XX  
 XX 01-DEC-1999 (first entry)  
 DT  
 XX Murine Helios-2 protein.  
 DE  
 XX Helios; Ikaros; Aiolos; DNA binding; transcription; activation;  
 KW immune disorder; proliferative disorder; leukaemia; Hodgkin's lymphoma;  
 KW T cell lymphoma.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 116..136  
 FT /label= zinc\_finger\_1  
 FT /function= "Involved in DNA binding"  
 FT Domain 144..164  
 FT /label= zinc\_finger\_2  
 FT /function= "Involved in DNA binding"  
 FT Domain 172..193  
 FT /label= zinc\_finger\_3  
 FT /function= "Involved in DNA binding"  
 FT Domain 260..350  
 FT /label= Transcription\_activation\_domain  
 FT Domain 447..468  
 FT /label= zinc\_finger\_4  
 FT /function= "Involved in dimerisation"  
 FT Domain 475..497  
 FT /label= zinc\_finger\_5  
 FT /function= "Involved in dimerisation"  
 FT  
 XX WO9943208-A1.  
 PN  
 PD 02-SEP-1999.  
 XX  
 PF 26-FEB-1999; 99WO-US04224.  
 XX  
 PR 27-FEB-1996; 98US-0076325.  
 XX  
 PA (GEO ) GEN HOSPITAL CORP.  
 XX

PI Georgopoulos K, Kelley C, Morgan BA;  
 XX  
 DR WPI; 1999-527521/44.  
 DR N-PSDB; AA211962.  
 XX  
 PT New nucleic acids encoding mouse Helios-1 and -2 polypeptides,  
 PT useful for treating T cell, B cell and immune disorders and  
 PT proliferative disorders  
 XX  
 PS Claim 7; Page 95-96; 107pp; English.  
 XX  
 CC This sequence represents a murine Helios-2 protein. Helios protein is  
 CC expressed in haematopoietic stem cells. It is related to both Aiolos  
 CC and Ikaros proteins and is a transcriptional activator of one or more  
 CC lymphoid genes as a homodimer or as a heterodimer with Aiolos or Ikaros  
 CC proteins. Murine Helios-2 protein comprises 5 zinc finger domains: the  
 CC 3 N-terminal zinc fingers are involved in binding DNA, and the 2  
 CC C-terminal zinc fingers mediate dimerisation. Helios-1 and -2 are the  
 CC products of alternative RNA splicing, with Helios-2 missing zinc finger  
 CC 1. Helios cDNAs were isolated from an Aiolos-mutant mouse spleen cDNA  
 CC library using degenerate primers encoding conserved N- and C-terminal  
 CC Ikaros zinc finger sequences. Full-length cDNAs were generated using the  
 CC RACE (rapid amplification of cDNA ends) technique using primers  
 CC AA211966- AA211969. Disorders associated with the inappropriate  
 CC expression of Helios genes include T-cell, B-cell and immune disorders  
 CC and proliferative disorders (such as leukaemia, Hodgkin's lymphoma and  
 CC cutaneous T cell lymphoma). Nucleic acids encoding Helios may be used to  
 CC identify similar nucleic sequences in biological samples and assess the  
 CC risk a patient has of developing a disorder associated with the  
 CC inappropriate expression of Helios genes. It may also be used for  
 CC treating those disorders by gene therapy protocols. Additionally, such  
 CC nucleotides may also be used in the manufacture of Helios proteins by  
 CC recombinant techniques.  
 XX  
 SQ Sequence 500 AA;

Query Match 77.6%; Score 45; DB 20; Length 500;  
 Best Local Similarity 77.8%; Pred. No. 8.6;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRCQ 9  
 Db 110 gprfhcncq 118  
 | |||| |

RESULT 3  
 AAY34146  
 ID AAY34146 standard; Protein; 526 AA.  
 AC AAY34146;  
 XX  
 XX 01-DEC-1999 (first entry)  
 DT  
 XX Murine Helios-1 protein.  
 DE  
 XX Helios; Ikaros; Aiolos; DNA binding; transcription; activation;  
 KW immune disorder; proliferative disorder; leukaemia; Hodgkin's lymphoma;  
 KW T cell lymphoma.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 114..134  
 FT /label= zinc\_finger\_1  
 FT /function= "Involved in DNA binding"  
 FT Domain 142..162  
 FT /label= zinc\_finger\_2  
 FT /function= "Involved in DNA binding"  
 FT Domain 170..190  
 FT /label= zinc\_finger\_3  
 FT /function= "Involved in DNA binding"  
 FT Domain 198..219



```

FT FT /label= zinc_finger_4
FT FT /function= "Involved in DNA binding"
FT FT 286..376
FT FT Domain
FT FT /label= Transcription_activation_domain
FT FT 473..494
FT FT /label= zinc_finger_5
FT FT /function= "Involved in dimerisation"
FT FT 501..523
FT FT /label= zinc_finger_6
FT FT /function= "Involved in dimerisation"
XX XX
XX PN WO9943208-A1.
XX XX
XX PD 02-SEP-1999.
XX XX
XX PF 26-FEB-1999; 99WO-US04224.
XX XX
XX PR 27-FEB-1998; 98US-0076325.
XX XX (GEO ) GEN HOSPITAL CORP.
XX PA
XX PI Georgopoulos K, Kelley C, Morgan BA;
XX XX
XX DR WPI; 1999-527521/44.
XX DR N-PSDB; AAY11960.
XX XX
XX PT New nucleic acids encoding mouse Helios-1 and -2 polypeptides,
XX PT useful for treating T cell, B cell and immune disorders and
XX PT proliferative disorders
XX XX
XX PS Claim 7; Page 92-93; 107pp; English.
XX XX
XX CC This sequence represents a murine Helios-1 protein. Helios protein is
XX CC expressed in haematopoietic stem cells. It is related to both Aiolos
XX CC and Ikaros proteins and is a transcriptional activator of one or more
XX CC lymphoid genes as a homodimer or as a heterodimer with Aiolos or Ikaros
XX CC proteins. Murine Helios-1 protein comprises 6 zinc finger domains: the
XX CC 4 N-terminal zinc fingers are involved in binding DNA, and the 2
XX CC C-terminal zinc fingers mediate dimerisation. Helios-1 and -2 are the
XX CC products of alternative RNA splicing, with Helios-2 missing zinc finger
XX CC 1. Helios cDNAs were isolated from an Aiolos-mutant mouse spleen cDNA
XX CC library using degenerate primers encoding conserved N- and C-terminal
XX CC Ikaros zinc finger sequences. Full-length cDNAs were generated using the
XX CC RACE (rapid amplification of cDNA ends) technique using primers
XX CC AAZ11966- AAZ11969. Disorders associated with the inappropriate
XX CC expression of Helios genes include T-cell, B-cell and immune disorders
XX CC and proliferative disorders (such as leukaemia, Hodgkin's lymphoma and
XX CC cutaneous T cell lymphoma). Nucleic acids encoding Helios may be used to
XX CC identify similar nucleic sequences in biological samples and assess the
XX CC risk a patient has of developing a disorder associated with the
XX CC inappropriate expression of Helios genes. It may also be used for
XX CC treating those disorders by gene therapy protocols. Additionally, such
XX CC nucleotides may also be used in the manufacture of Helios proteins by
XX CC recombinant techniques.
XX SQ Sequence 526 AA;

Query Match 77.6%; Score 45; DB 20; Length 526;
Best Local Similarity 77.8%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCQ 9
Db 136 gerpfhcnq 144
| | | | | | |
| | | | | | |

RESULT 4
AAY34148
ID AAY34148 standard; Protein; 526 AA.
XX
XX AC AAY34148;
XX XX

```

```

DT 01-DEC-1999 (first entry)
XX XX Human Helios-2 protein.
XX XX
XX KW Helios; Ikaros; Aiolos; DNA binding; transcription; activation;
XX KW immune disorder; proliferative disorder; leukaemia; Hodgkin's lymphoma;
XX KW T cell lymphoma; ds.
XX XX
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
FT FT Domain II4..134
FT FT /label= zinc_finger_1
FT FT /function= "Involved in DNA binding"
FT FT Domain 142..162
FT FT /label= zinc_finger_2
FT FT /function= "Involved in DNA binding"
FT FT Domain 170..190
FT FT /label= zinc_finger_3
FT FT /function= "Involved in DNA binding"
FT FT Domain 198..219
FT FT /label= zinc_finger_4
FT FT /function= "Involved in DNA binding"
FT FT Domain 286..376
FT FT /label= Transcription_activation_domain
FT FT Domain 473..494
FT FT /label= zinc_finger_5
FT FT /function= "Involved in dimerisation"
FT FT Domain 501..523
FT FT /label= zinc_finger_6
FT FT /function= "Involved in dimerisation"
XX XX
XX PN WO9943208-A1.
XX XX
XX PD 02-SEP-1999.
XX XX
XX PF 26-FEB-1999; 99WO-US04224.
XX XX
XX PR 27-FEB-1998; 98US-0076325.
XX XX (GEO ) GEN HOSPITAL CORP.
XX XX
XX PI Georgopoulos K, Kelley C, Morgan BA;
XX XX WPI; 1999-527521/44.
XX DR N-PSDB; AAY34148.
XX XX
XX PT New nucleic acids encoding mouse Helios-1 and -2 polypeptides,
XX PT useful for treating T cell, B cell and immune disorders and
XX PT proliferative disorders
XX XX
XX PS Claim 7; Page 99; 107pp; English.
XX XX
XX CC This sequence represents a human Helios-2 protein. Helios protein is
XX CC expressed in haematopoietic stem cells. It is related to both Aiolos
XX CC and Ikaros proteins and is a transcriptional activator of one or more
XX CC lymphoid genes as a homodimer or as a heterodimer with Aiolos or Ikaros
XX CC proteins. Human Helios-2 protein comprises 6 zinc finger domains: the
XX CC 4 N-terminal zinc fingers are involved in binding DNA, and the 2
XX CC C-terminal zinc fingers mediate dimerisation. Helios cDNAs were
XX CC isolated using degenerate primers derived from the conserved N- and
XX CC C-terminal zinc finger sequences of murine Ikaros. Full-length cDNAs
XX CC were generated using the RACE (rapid amplification of cDNA ends)
XX CC technique. Disorders associated with the inappropriate expression of
XX CC Helios genes include T-cell, B-cell and immune disorders and
XX CC proliferative disorders (such as leukaemia, Hodgkin's lymphoma and
XX CC cutaneous T cell lymphoma). Nucleic acids encoding Helios may be used to
XX CC identify similar nucleic sequences in biological samples and assess the
XX CC risk a patient has of developing a disorder associated with the
XX CC inappropriate expression of Helios genes. It may also be used for
XX CC treating those disorders by gene therapy protocols.
XX CC Additionally, such nucleotides may also be used in the manufacture of
XX CC Helios proteins by recombinant techniques.

```

```
XX SQ Sequence 526 AA;
Query Match 77.6%; Score 45; DB 20; Length 526;
Best Local Similarity 77.8%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPRPFHCRCQ 9
| | | | |
Db 136 gerpfhcnq 144

RESULT 5
AAY87745
ID AAY87745 standard; Protein; 533 AA.
XX
AC AAY87745;
XX
DT 21-AUG-2000 (first entry)
XX
DE Murine embryonic brain Eos protein.
XX
KW Eos gene; drug development; cerebral nervous disease; gene therapy;
murine.
XX
OS Mus sp.
XX
PN JP2000093177-A.
XX
PD 04-APR-2000.
XX
PF 18-SEP-1998; 98JP-0264192.
XX
PR 18-SEP-1998; 98JP-0264192.
XX
PA (SAKA ) OTSUKA PHARM CO LTD.
XX
WPI; 2000-378971/33.
DR N-PSDB; AAA12375.
XX
XX Isolated Eos gene useful for drug development for various nervous
diseases -
XX
Claim 1a; Page 22-23; 29pp; Japanese.
XX
This invention describes a novel isolated Eos protein and its
encoding nucleic acid. The Eos gene can be used for the development
of new drugs for various cerebral nervous diseases. Independent claims
are also included for the following: (1) a cloned cDNA carrying the
above Eos gene and exhibiting an activity of inhibiting formation of
abnormality in nervous system; (2) an Eos gene expression product having
the above amino acid sequence; (3) a cloned cDNA expressing Eos protein;
its fragment, its derivative and its homolog; (4) a human homolog of Eos
gene; (5) an expression product of the human homolog of Eos gene; (6)
inhibiting formation of abnormality in nervous system including a step of
introducing Eos gene in a cell by using the above cloned cDNA; (7) a drug
composition, that is a gene therapy agent, containing the above Eos gene
as the active component. This represents the murine embryonic brain Eos
protein described in the method of the invention.
XX SQ Sequence 533 AA;
Query Match 77.6%; Score 45; DB 21; Length 533;
Best Local Similarity 77.8%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPRPFHCRCQ 9
| | | | |
Db 130 gerpfhcnq 138

RESULT 6
AAB95116
ID AAB95116 standard; Protein; 725 AA.
XX
AC AAB95116;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17104.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
XX
Claim 8; SEQ ID 17104; 2537pp + CD ROM; English.
XX
The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.
XX SQ Sequence 725 AA;
Query Match 77.6%; Score 45; DB 22; Length 725;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPRPFHCRCQ 9
| | | | |
Db 547 gprphctq 555
```

```

RESULT 7
AAV01161
ID AAY01161 standard; Protein; 19 AA.
XX
AC AAY01161;
XX
DT 18-MAY-1999 (first entry)
XX
DE Polypeptide fragment encoded by gene 1.
XX
KW Human; secreted protein; gene therapy; protein therapy; tissue; cancer;
KW tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;
KW developmental abnormality; foetal deficiency; Alzheimer's disease;
KW cognitive disorder; schizophrenia; immunological disorder; mood disorder;
KW immune deficiency disease; respiratory disorder; arthritis; skeletal;
KW haematopoietic disorder; neural; osteoporosis; metabolic disorders;
KW cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.
XX
OS Homo sapiens.
XX
PN WO9901020-A2.
XX
PD 14-JAN-1999.
XX
PF 30-JUN-1998; 98WO-US13608.
XX
PR 12-SEP-1997; 97US-0058663.
PR 01-JUL-1997; 97US-0051381.
PR 01-JUL-1997; 97US-0051480.
PR 12-SEP-1997; 97US-0058598.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Endress GA, Feng P, Rosen CA, Ruben SM;
XX
WPI: 1999-105683/09.
DR N-PSDB; AAX22111.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune deficiency diseases or blood
PT disorders
XX
PS Disclosure; Page 6; 179pp; English.
XX
CC The invention relates to nucleic acid sequences (AAX22111 to AAX22134)
CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted
CC protein gene sequences are deposited with the ATCC under deposit number
CC ATCC 209118. Host cells comprising recombinant vectors containing the
CC nucleic acid sequences are used for the recombinant production of the
CC secreted proteins. The polynucleotide and amino acid sequences are useful
CC for are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Pathological conditions can
CC be also diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Specific uses are described for each of the
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, developmental abnormalities and foetal deficiencies,
CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,
CC schizophrenia, immunological disorders, immune deficiency diseases
CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,
CC haematopoietic disorders, neural disorders, skeletal disorders,
CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine
CC disorders or gastrointestinal disorders. The polypeptides are also useful
CC for identifying their binding partners. The present sequence represents a
CC polypeptide fragment encoded by a gene of the invention (see descriptor
CC line for gene number).
XX
SQ Sequence 19 AA;

```

Query Match 74.1%; Score 43; DB 20; Length 19;  
Best Local Similarity 77.8%; Pred. No. 0.83;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRCQ 9  
| | | | | | |  
Db 10 gerpfhcdq 18

RESULT 8  
AAV01159  
ID AAY01159 standard; Protein; 259 AA.  
XX  
AC AAY01159;  
XX  
DT 18-MAY-1999 (first entry)  
XX  
DE Polypeptide fragment encoded by gene 1.  
XX  
KW Human; secreted protein; gene therapy; protein therapy; tissue; cancer;  
KW tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;  
KW developmental abnormality; foetal deficiency; Alzheimer's disease;  
KW cognitive disorder; schizophrenia; immunological disorder; mood disorder;  
KW immune deficiency disease; respiratory disorder; arthritis; skeletal;  
KW haematopoietic disorder; neural; osteoporosis; metabolic disorders;  
KW cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WO9901020-A2.  
XX  
PD 14-JAN-1999.  
XX  
PF 30-JUN-1998; 98WO-US13608.  
XX  
PR 12-SEP-1997; 97US-0058663.  
PR 01-JUL-1997; 97US-0051381.  
PR 01-JUL-1997; 97US-0051480.  
PR 12-SEP-1997; 97US-0058598.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Carter KC, Endress GA, Feng P, Rosen CA, Ruben SM;  
XX  
WPI: 1999-105683/09.  
DR N-PSDB; AAX22111.  
XX  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune deficiency diseases or blood  
PT disorders  
XX  
PS Disclosure; Page 6; 179pp; English.  
XX  
CC The invention relates to nucleic acid sequences (AAX22111 to AAX22134)  
CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted  
CC protein gene sequences are deposited with the ATCC under deposit number  
CC ATCC 209118. Host cells comprising recombinant vectors containing the  
CC nucleic acid sequences are used for the recombinant production of the  
CC secreted proteins. The polynucleotide and amino acid sequences are useful  
CC for are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Pathological conditions can  
CC be also diagnosed by determining the amount of the new polypeptides in a  
CC sample or by determining the presence of mutations in the new  
CC polynucleotides. Specific uses are described for each of the  
CC in, and include developing products for the diagnosis or treatment of  
CC cancer, tumours, developmental abnormalities and foetal deficiencies,  
CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,  
CC schizophrenia, immunological disorders, immune deficiency diseases  
CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,  
CC haematopoietic disorders, neural disorders, skeletal disorders,  
CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine  
CC disorders or gastrointestinal disorders. The polypeptides are also useful  
CC for identifying their binding partners. The present sequence represents a  
CC polypeptide fragment encoded by a gene of the invention (see descriptor  
CC line for gene number).

CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine  
 CC disorders or gastrointestinal disorders. The polypeptides are also useful  
 CC for identifying their binding partners. The present sequence represents a  
 CC polypeptide fragment encoded by a gene of the invention (see descriptor  
 CC line for gene number).

XX Sequence 259 AA;

Query Match 74.1%; Score 43; DB 20; Length 259;  
 Best Local Similarity 77.8%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRC 9  
 I I I I I I I  
 Db 65 gerpfhcdq 73

RESULT 9  
 AAY73354  
 ID AAY73354 standard; Protein; 259 AA.

XX AC AAY73354;

XX DT 24-FEB-2000 (first entry)

XX DE HTRM clone 1910984 protein sequence.

XX KW HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
 KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;  
 KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
 KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.

XX OS Homo sapiens.

XX PN WO9957144-A2.

XX PD 11-NOV-1999.

XX PF 04-MAY-1999; 99WO-US09935.

XX PR 05-MAY-1998; 98US-0084254.

XX PR 07-AUG-1998; 98US-0095827.

XX PR 02-OCT-1998; 98US-0102745.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;

XX PI Gerstein EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;

XX DR WPI; 2000-052941/04.

XX DR N-PSDB; AA252439.

XX PT New peptides useful for diagnosis, prevention and treatment of cancer

XX PT and immune disorders -

XX PS Claim 1; Page 115-116; 193pp; English.

XX CC AAY73325-V73389 are human transcriptional regulator molecule (HTRM)  
 CC protein sequences. The HTRM protein and nucleotide sequences are useful  
 CC for preventing or treating disorders associated with decreased expression  
 CC or activity of HTRM which include cell proliferative disorders such as  
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
 CC leukaemia; immune disorders such as AIDS, Addison's disease, diabetes  
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
 CC of the HTRM polypeptides are useful for treating or preventing disorders  
 CC associated with increased expression or activity of HTRMs. HTRM  
 CC polypeptides, their immunogenic fragments or oligopeptides are useful  
 CC for screening libraries of compounds in drug screening techniques.  
 CC Polynucleotides encoding HTRM are useful for blocking the transcription  
 CC of mRNA and regulating gene function by modulating the activity of  
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or

CC treat disorder associated with decreased HTRM expression. Antibodies  
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
 CC for diagnosing disorders associated with the expression of HTRM,  
 CC particularly in assays that detect the expression of HTRM. Nucleotide  
 CC sequences encoding HTRM may be useful to generate hybridization probes  
 CC useful in mapping the naturally occurring genomic sequence and to detect  
 CC differences in gene sequences among normal, carrier and affected  
 CC individuals. Using diagnostic assays, cancer can be detected prior to  
 CC the appearance of clinical symptoms and thereby progression of cancer can  
 CC be prevented by aggressive treatment or preventive measures.

XX Sequence 259 AA;

Query Match 74.1%; Score 43; DB 21; Length 259;  
 Best Local Similarity 77.8%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRC 9  
 I I I I I I I  
 Db 65 gerpfhcdq 73

RESULT 10  
 AAU15915  
 ID AAU15915 standard; Protein; 121 AA.

XX AC AAU15915;

XX DT 07-NOV-2001 (first entry)

XX DE Human novel secreted protein, Seq ID 868.

XX KW Human; immunosuppressive; antiarthritic; antirheumatic;

XX KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;

XX KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

XX KW vulnary; secreted protein; rheumatoid arthritis;

XX KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

XX KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;

XX KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;

XX KW corneal infection; wound healing; epithelial cell proliferation;

XX KW skin ageing; food additive; preservative; antiproliferative.

XX OS Homo sapiens.

XX PN WO200155322-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01341.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUL-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225271.  
PR 14-AUG-2000; 2000US-0225275.  
PR 14-AUG-2000; 2000US-0225278.  
PR 14-AUG-2000; 2000US-0225279.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234297.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-488783/53.  
N-PSDB; AAS25902.

New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives -  
Claim 11; SEQ ID No 868; 980pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before

CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention.

Query Match 72.4%; Score 42; DB 22; Length 121;  
 Best Local Similarity 77.8%; Pred. No. 7.2;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRCQ 9  
 | | | | |  
 Db 90 gerpfqcrq 98

RESULT 11  
 AAU28084  
 ID AAU28084 standard; Protein: 529 AA.

XX AAU28084;

DT 18-DEC-2001 (first entry)

DE Novel human secretory protein, Seq ID No 253.

XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
 KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
 KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
 KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
 KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
 KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
 KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
 KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
 KW fertility; analgesic; pain; antigen.

XX Homo sapiens.

OS WO200166689-A2.

PN 13-SEP-2001.

XX 05-MAR-2001; 2001WO-US04942.

XX 07-MAR-2000; 2000US-0519705.

PR 19-MAY-2000; 2000US-0574454.

PR 17-JUN-2000; 2000US-0596193.

PR 14-JUL-2000; 2000US-0616847.

PR 19-SEP-2000; 2000US-0665363.

PR 20-OCT-2000; 2000US-0693267.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;

PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;

XX WPI: 2001-589934/66.

DR N-PSDB; AAS44984.

XX Novel polypeptides and nucleic acids obtained from cDNA libraries

PT prepared from various human tissues, for diagnosis and treatment of

PT cancer, neurological, inflammatory, and autoimmune disorders -

PS Example 4; SEQ ID No 253; 107pp; English.

XX The invention relates to novel isolated human secreted polypeptides (I)

CC and polynucleotides (II). (I) and (II) are useful for treating

CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,

CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is

CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (I) induces the proliferation of neural cells and regeneration of nerve  
 CC brain tissue and is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth, and in tissue repair, healing of burns, incisions,  
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
 CC disorders, or periodontal disease. Furthermore, (I) is also useful for  
 CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, various immune deficiencies and  
 CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 CC reactions and conditions, such as asthma or other respiratory problems.  
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 CC analgesic effects or other pain reducing effects, immunoglobulin like  
 CC activity and can act as an antigen in a vaccine composition to raise an  
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
 CC amino acid sequences of the invention.

SQ Sequence 529 AA;

Query Match 72.4%; Score 42; DB 22; Length 529;

Best Local Similarity 77.8%; Pred. No. 30;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPRPFHCRCQ 9

| | | | |

Db 498 gerpfqcrq 506

RESULT 12

AAU61505

ID AAU61505 standard; Protein: 162 AA.

XX AAU61505;

XX 27-FEB-2002 (first entry)

DT Propionibacterium acnes immunogenic protein #22401.

DE SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.

DR N-PSDB; AAS59618.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 22700; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 162 AA;

Query Match 70.7%; Score 41; DB 22; Length 162;

Best Local Similarity 66.7%; Pred. NO. 14;

Matches 8; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GRPFH----CR 8

DB 94 GRPFHVLlrcr 105

RESULT 13

AAU50534

ID AAU50534 standard; Protein; 53 AA.

AC AAU50534;

DT 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #11430.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AA559549.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX

PS Example 1; SEQ ID No 11729; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 53 AA;

Query Match 69.0%; Score 40; DB 22; Length 53;

Best Local Similarity 85.7%; Pred. No. 7.2;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRPFHCR 8

DB 28 prpfhcr 34

RESULT 14

AAG03279

ID AAG03279 standard; Protein; 76 AA.

AC AAG03279;

DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7360.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PSDB; AAC03285.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 7360; 71pp + CD-ROM; English.  
PS  
XX The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.  
XX  
SQ Sequence 76 AA;

Query Match 69.0%; Score 40; DB 21; Length 76;  
Best Local Similarity 66.7%; Pred. No. 10;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPRPHCRQ 9  
||| |||:  
Db 4 gprahcrr 12

RESULT 15  
AAU16426  
ID AAU16426 standard; Protein; 140 AA.  
XX  
AC AAU16426;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human novel secreted protein, Seq ID 1379.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO200153322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.

26-JUL-2000; 2000US-0220963.  
26-JUL-2000; 2000US-0220964.  
14-AUG-2000; 2000US-0224518.  
14-AUG-2000; 2000US-0224519.  
14-AUG-2000; 2000US-0225213.  
14-AUG-2000; 2000US-0225214.  
14-AUG-2000; 2000US-0225266.  
14-AUG-2000; 2000US-0225267.  
14-AUG-2000; 2000US-0225268.  
14-AUG-2000; 2000US-0225270.  
14-AUG-2000; 2000US-0225447.  
14-AUG-2000; 2000US-0225757.  
14-AUG-2000; 2000US-0225758.  
14-AUG-2000; 2000US-0225759.  
18-AUG-2000; 2000US-0226279.  
22-AUG-2000; 2000US-0226681.  
22-AUG-2000; 2000US-0226868.  
22-AUG-2000; 2000US-0227182.  
23-AUG-2000; 2000US-0227009.  
30-AUG-2000; 2000US-0228924.  
01-SEP-2000; 2000US-0229287.  
01-SEP-2000; 2000US-0229343.  
01-SEP-2000; 2000US-0229344.  
01-SEP-2000; 2000US-0229345.  
05-SEP-2000; 2000US-0229509.  
05-SEP-2000; 2000US-0229513.  
06-SEP-2000; 2000US-0230437.  
06-SEP-2000; 2000US-0230438.  
08-SEP-2000; 2000US-0231242.  
08-SEP-2000; 2000US-0231243.  
08-SEP-2000; 2000US-0231244.  
08-SEP-2000; 2000US-0231413.  
08-SEP-2000; 2000US-0231414.  
08-SEP-2000; 2000US-0232080.  
08-SEP-2000; 2000US-0232081.  
12-SEP-2000; 2000US-0231968.  
14-SEP-2000; 2000US-0232397.  
14-SEP-2000; 2000US-0232398.  
14-SEP-2000; 2000US-0232399.  
14-SEP-2000; 2000US-0232400.  
14-SEP-2000; 2000US-0232401.  
14-SEP-2000; 2000US-0233063.  
14-SEP-2000; 2000US-0233064.  
21-SEP-2000; 2000US-0233065.  
21-SEP-2000; 2000US-0234223.  
21-SEP-2000; 2000US-0234274.  
25-SEP-2000; 2000US-0234997.  
25-SEP-2000; 2000US-0234998.  
26-SEP-2000; 2000US-0235484.  
27-SEP-2000; 2000US-0235834.  
27-SEP-2000; 2000US-0235836.  
29-SEP-2000; 2000US-0236327.  
29-SEP-2000; 2000US-0236367.  
29-SEP-2000; 2000US-0236368.  
29-SEP-2000; 2000US-0236369.  
29-SEP-2000; 2000US-0236370.  
02-OCT-2000; 2000US-0236802.  
02-OCT-2000; 2000US-0237037.  
02-OCT-2000; 2000US-0237038.  
02-OCT-2000; 2000US-0237039.  
02-OCT-2000; 2000US-0237040.  
13-OCT-2000; 2000US-0239935.  
13-OCT-2000; 2000US-0239937.  
20-OCT-2000; 2000US-0240960.  
20-OCT-2000; 2000US-0241221.  
20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241786.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
01-NOV-2000; 2000US-0244617.  
08-NOV-2000; 2000US-0246474.



PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

WPI: 2001-488783/53.

DR N-PSDB; AAS26413.

XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

XX Claim 11; SEQ ID NO 1379; 980pp; English.

XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Query Match 69.0%; Score 40; DB 22; Length 140;

Best Local Similarity 85.7%; Pred. No. 18;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPRPHHC 7

Db. 58 gerpfhc 64

Search completed: May 24, 2002, 16:48:04

Job time: 244 sec

THIS PAGE INTENTIONALLY LEFT BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:02:51 ; Search time 107.9 seconds  
(without alignments)  
46.495 Million cell updates/sec

Title: US-09-730-379E-4

Perfect score: 135

Sequence: 1 IGSVYRLPPLRKGEVLPLPEANFPSPFLP 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvivirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	77.4	525	11 Q99PS6	Q99ps6 mus musculus
2	120	77.4	525	11 Q9ESB3	Q9esb3 mus musculus
3	120	77.4	525	11 Q99PS5	Q99ps5 mus musculus
4	120	77.4	546	6 Q9BGU1	Q9bgul bos taurus
5	114	73.5	515	11 Q99PS7	Q99ps7 rattus norv
6	108	69.7	510	11 Q9ESB2	Q9esb2 rattus norv
7	108	69.7	525	11 Q99PS8	Q99ps8 rattus norv
8	82	40.0	891	16 Q9RWY5	Q9rwy5 deinococcus
9	58.5	37.7	117	16 Q9ZKX0	Q9zkx0 helicobacte
10	58.5	37.7	1154	16 Q9ZK97	Q9zk97 helicobacte
11	57.5	37.1	162	10 Q43516	Q43516 lycopersico
12	56.5	36.5	544	5 Q9GQC4	Q9gqc4 bombyx mori
13	56	36.1	1246	15 Q36977	O36977 walleye epi
14	56	36.1	1246	15 Q9WHJ2	Q9whj2 walleye epi
15	55.5	35.8	565	5 Q9GPG9	Q9gpg9 bombyx mand
16	54.5	35.2	967	4 Q96Q00	Q96q00 homo sapien

17	54.5	35.2	1003	4	Q96DN6	Q96dn6 homo sapien
18	54.5	35.2	1246	3	O15052	O15052 homo sapien
19	54.5	35.2	1464	3	O9C229	O9c229 neurospora
20	54.5	35.2	2382	4	Q9H4A3	Q9h4a3 homo sapien
21	54	34.8	625	2	Q9ED94	Q9ed94 azospirillum
22	53.5	34.5	405	5	O17087	O17087 caenorhabdi
23	53.5	34.5	494	13	Q9PU55	Q9pu55 gallus gall
24	53.5	34.5	533	11	Q922M0	Q922m0 mus musculu
25	53.5	34.5	543	5	Q9GV05	Q9gv05 bombyx mori
26	53.5	34.5	565	5	P90710	P90710 bombyx mori
27	53	34.2	486	5	Q9U1W1	Q9u1w1 caenorhabdi
28	53	34.2	599	10	P93797	P93797 volvox cart
29	53	34.2	884	5	Q9GU16	Q9gu16 caenorhabdi
30	53	34.2	1611	3	O42854	O42854 schizosacch
31	52.5	33.9	273	10	O9ARY7	O9ary7 oryza sativ
32	51.5	33.2	543	5	Q9GR93	Q9gr93 bombyx mori
33	51.5	33.2	552	5	Q9GV44	Q9gv44 spodoptera
34	51	32.9	258	5	Q9U191	Q9u191 leishmania
35	51	32.9	361	5	Q21847	Q21847 caenorhabdi
36	51	32.9	491	5	Q96OV9	Q96ov9 drosophila
37	51	32.9	534	2	O9KZE3	O9kze3 streptomyce
38	51	32.9	534	2	O86874	O86874 streptomyce
39	51	32.9	569	5	Q9VL56	Q9vl56 drosophila
40	51	32.9	703	10	Q9LHJ0	Q9lhj0 arabidopsis
41	50.5	32.6	875	3	Q9GWP8	Q9gwp8 aspergillus
42	50	32.3	170	2	Q9FBK2	Q9fbk2 streptomyce
43	50	32.3	488	10	Q9LK94	Q9lk94 arabidopsis
44	50	32.3	516	16	O07784	O07784 mycobacteri
45	50	32.3	796	2	O69908	O69908 streptomyce

#### ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	525 AA.
Q99PS6	AC Q99PS6;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)		
DE	HISTIDINE-RICH GLYCOPROTEIN.		
GN	MMHRG.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BALB/C; TISSUE=LIVER;		
RA	Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,		
RA	Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;		
RT	"Molecular diversity of mammalian histidine-rich glycoprotein.";		
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB055897; BAB33094.1; -		
DR	InterPro: IPR000010; Cystatin.		
DR	Pfam: PF00031; cystatin; 1.		
DR	SMART: SM00043; Cy; 2.		
SQ	SEQUENCE 525 AA; 59090 MW; A83E93A439CFB3AC CRC64;		

Query Match	77.4%;	Score 120;	DB 11;	Length 525;
Best Local Similarity	82.8%;	Pred. No. 6.5e-09;		
Matches 24;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	1	IGSVYRLPPLRKGEVLPLPEANFPSPFLP 29		
Db	455	IGVYRLPPLNICEVLTLPEANFPSPFLP 483		

RESULT 2	PRELIMINARY;	PRT;	525 AA.
Q9ESB3	ID Q9ESB3		

AC Q9EB3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN.  
GN HRG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RA MEDLINE=20307726; PubMed=10849117;  
RX Hulet M.D., Parish C.R.;  
RT "Murine histidine-rich glycoprotein: cloning, characterization and  
cellular origin";  
RL Immunol. Cell Biol. 78:280-287(2000).  
DR EMBL; AF194028; AAG28416.1; -;  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
SQ SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;

Query Match 77.4%; Score 120; DB 11; Length 525;  
Best Local Similarity 82.8%; Pred. No. 6.5e-09;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IGSVRLPPLRKGEVLPPEANFPSPPLP 29  
II IIIIIII IIII IIIIIIIII II  
DB 455 IGVYRLPPLNIGEVLTLPPEANFPSPSLP 483

RESULT 3  
Q99PS5  
ID Q99PS5 PRELIMINARY; PRT; 525 AA.  
AC Q99PS5;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).  
GN MHRG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tsuchida N., Wakabayashi S., Jahnhen-Dechent W., Koide T.;  
RT "Structure of mouse histidine-rich glycoprotein gene";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB055898; BAB33095.1; -;  
DR EMBL; BC011168; AAH11168.1; -;  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match 77.4%; Score 120; DB 11; Length 525;  
Best Local Similarity 82.8%; Pred. No. 6.5e-09;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IGSVRLPPLRKGEVLPPEANFPSPPLP 29  
II IIIIIII IIII IIIIIIIII II  
DB 455 IGVYRLPPLNIGEVLTLPPEANFPSPSLP 483

RESULT 4  
Q9BGU1  
ID Q9BGU1 PRELIMINARY; PRT; 546 AA.  
AC Q9BGU1;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN.  
GN BTHRG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
RT "Molecular diversity of mammalian histidine-rich glycoprotein";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB055894; BAB33091.1; -;  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
SQ SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 77.4%; Score 120; DB 6; Length 546;  
Best Local Similarity 78.6%; Pred. No. 6.8e-09;  
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 GSVYRLPPLRKGEVLPPEANFPSPPLP 29  
I :IIII :IIIIIIIIIIII II  
DB 477 GYIHLPLSLKKEVLPPEANFPSPSLP 504

RESULT 5  
Q99PS7  
ID Q99PS7 PRELIMINARY; PRT; 515 AA.  
AC Q99PS7;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE HISTIDINE-RICH GLYCOPROTEIN 2.  
GN RHRG2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,  
Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;  
RT "Molecular diversity of mammalian histidine-rich glycoprotein";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB055896; BAB33093.1; -;  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 73.5%; Score 114; DB 11; Length 515;  
Best Local Similarity 79.3%; Pred. No. 4.6e-08;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 IGSVYRLPPLRKGEVLPPEANFPSPPLP 29  
II IIIIIII IIII IIIIIIIII II  
DB 445 IGVYRLPPLNIGEVLTLPPEANFPSPSLP 473

```

RESULT 6
Q9ESB2
ID Q9ESB2 PRELIMINARY; PRT; 510 AA.
AC Q9ESB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN.
HRG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEWIS;
RX MEDLINE=20307726; PubMed=10849117;
RA Hueltt M.D., Farish C.R.;
RT "Murine histidine-rich glycoprotein: cloning, characterization and
RT cellular origin.";
RL Immunol. Cell Biol. 78:280-287(2000).
DR EMBL: AF194029; AAG28417.1; -.
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
DR SMART: SM00043; CY; 2.
DR SMART: SM00043; CY; 2.
SQ SEQUENCE 510 AA; 57581 MW; 508E6E06AA2ED58E CRC64;

Query Match 69.7%; Score 108; DB 11; Length 510;
Best Local Similarity 75.9%; Pred. No. 3.3e-07;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 IGSVYRLPLRKGEVLPLPEANFPSPPLP 29
II IIIIIII IIII IIIIIII II
DB 440 IGVYRLPLNVGVLTPPEANFPIFSLP 468

RESULT 7
Q99PS8
ID Q99PS8 PRELIMINARY; PRT; 525 AA.
AC Q99PS8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN 1.
GN RNHRL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Onishi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB055895; BAB33092.1; -.
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
DR SMART: SM00043; CY; 2.
DR SMART: SM00043; CY; 2.
SQ SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match 69.7%; Score 108; DB 11; Length 525;
Best Local Similarity 75.9%; Pred. No. 3.4e-07;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 IGSVYRLPLRKGEVLPLPEANFPSPPLP 29
II IIIIIII IIII IIIIIII II
DB 455 IGVYRLPLNVGVLTPPEANFPIFSLP 483

```

```

RESULT 8
Q9RWY5
ID Q9RWY5 PRELIMINARY; PRT; 891 AA.
AC Q9RWY5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 93.6 KDA PROTEIN.
GN DR0530.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AF001911; AAF10112.1; -.
DR TIGR: DR0530; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 891 AA; 93612 MW; DC5B0F59332180F1 CRC64;

Query Match 40.0%; Score 62; DB 16; Length 891;
Best Local Similarity 56.5%; Pred. No. 2.1;
Matches 13; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 7 LPPLRKGEVLPLPEANFPSPPLP 29
III III III I III
DB 366 LPPVANGEVAPLPVTPVPVPLP 388

RESULT 9
Q9ZKX0
ID Q9ZKX0 PRELIMINARY; PRT; 117 AA.
AC Q9ZKX0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PUTATIVE.
GN JHP0814.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL: AE001510; AAD06384.1; -.
KW Complete proteome.
SQ SEQUENCE 117 AA; 13400 MW; D2A1DCE21BADBF80 CRC64;

Query Match 37.7%; Score 58.5; DB 16; Length 117;
Best Local Similarity 40.5%; Pred. No. 0.76;
Matches 15; Conservative 3; Mismatches 8; Indels 11; Gaps 2;

```

1:

Db 126 OLPLPKPDLPMPOLPKPOLPLP 15

DE CHITINASE.  
QS Bombyx mori (silk moth).

```

RN      [1]
RP      SEQUENCE FROM N.A.
RP      STRAIN=SHUNREI X SUBGETSU:
RC      MEDLINE=20545518; PubMed=10944482;
RX      Mikitani K., Sugasaki T., Shimada T., Kobayashi M., Gustafsson J.A.,
RT      "The Chinese Gene of the Silkworm, Bombyx mori, Contains a Novel T-
RL      like Transposable Element".
RL      J. Biol. Chem. 275:37725-37732(2000)

```

CC -|- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL: AF273695; AAG37105.1; -.

2:

DR SMARI, SM004394, CHCDD2, 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SO SEQUENCE 544 AA; 61069 MW; 8987

1;

GN POL.

RT "Two closely related but distinct retroviruses are associated with

RL J. Virol. 72:3484-3490(1998).  
DR EMBL; AF014793; AAC59311.1; -



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2002, 17:03:27 ; Search time 31.45 Seconds  
(without alignments)  
35.703 Million cell updates/sec

Title: US-09-730-379E-4

Perfect score: 155

Sequence: 1 IGSVYRLPLRKGEVLPEANFPSEPLP 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155	100.0	525	1 HRG_HUMAN	P04196 homo sapien
2	120	77.4	396	1 HRG_BOVIN	P33433 bos taurus
3	120	77.4	326	1 HRG_RABIT	Q28640 oryctolagus
4	66.5	42.9	360	1 NO44_SOYBN	P04672 glycine max
5	54	34.8	428	1 FUCP_HAEIN	P44776 haemophilus
6	54	34.8	624	1 NIFA_AZOLI	P54929 azospirillum
7	53.5	34.5	533	1 LCP2_MOUSE	Q60787 mus musculus
8	51	32.9	511	1 MVIN_ECOLI	P75932 escherichia
9	50.5	32.6	461	1 TRIA_PIG	P50555 sus scrofa
10	50.5	32.6	875	1 XLNR_ASPNG	O42804 aspergillus
11	50	32.3	524	1 MVIN_SALTY	P37169 salmonella
12	50	32.3	1016	1 PMPH_CHLTR	O84880 chlamydia t
13	49.5	31.9	990	1 T3RE_SALTY	P40815 salmonella
14	49.5	31.9	3148	1 HD_FUGRU	P51112 fuigu rubrip
15	49	31.6	241	1 LAT_RAT	O70601 rattus norv
16	49	31.6	242	1 LAT_MOUSE	O54957 mus musculus
17	49	31.6	418	1 YD91_METJA	Q58786 methanococc
18	49	31.6	481	1 DOK1_HUMAN	O99704 homo sapien
19	49	31.6	3149	1 TEGU_EBV	P03186 epstein-bar
20	48	31.0	288	1 YV78_CAUCR	P58114 caulobacter
21	48	31.0	506	1 POL2_ARMV	P24819 arabis mosa
22	48	31.0	799	1 AFSK_STRCO	P54741 streptomyce
23	48	31.0	799	1 FLO8_YEAST	P40068 saccharomyc
24	48	31.0	980	1 PMPH_CHLMU	Q9pl44 chlamydia m
25	47.5	30.6	364	1 Y174_HUMAN	P53990 homo sapien
26	47.5	30.6	557	1 MK04_HUMAN	P31152 homo sapien
27	47	30.3	168	1 MB12_MYCTU	O05822 mycobacteri
28	47	30.3	488	1 RN23_MOUSE	Q9esn2 mus musculus
29	47	30.3	513	1 PDI_HORVU	P80284 hordeum vul
30	47	30.3	515	1 PDI_WHEAT	P52589 triticum ae
31	47	30.3	518	1 RN23_HUMAN	Q9hcm9 homo sapien
32	47	30.3	671	1 AMOI_ASPNG	P12556 aspergillus
33	47	30.3	951	1 HEX_ADE05	P04133 human adeno

34	47	30.3	1122	1 YG3C_YEAST	P53280 saccharomyc
35	46.5	30.0	1776	1 POLR_OYMV	P20127 ononis yell
36	46.5	30.0	2318	1 NTC3_MOUSE	Q61982 mus musculus
37	46	29.7	375	1 YHCG_ECOLI	P45423 escherichia
38	46	29.7	445	1 HPPD_ARATH	P93836 arabidopsis
39	46	29.7	556	1 PPBD_BACSU	P42251 bacillus su
40	46	29.7	640	1 ELL2_HUMAN	O00472 homo sapien
41	46	29.7	780	1 SAV_SULAC	Q07590 sulfolobus
42	45.5	29.4	533	1 LCP2_HUMAN	Q13094 homo sapien
43	45.5	29.4	1395	1 IFAG_HUMAN	Q04637 homo sapien
44	45.5	29.4	1453	1 MLH3_HUMAN	Q9uhcl1 homo sapien
45	45	29.0	367	1 A2HS_HUMAN	P02765 homo sapien

#### ALIGNMENTS

RESULT	1
HRG_HUMAN	
ID	HRG_HUMAN
AC	P04196;
DT	20-MAR-1987 (Rel. 04, Created)
DT	20-MAR-1987 (Rel. 04, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).
DE	glycoprotein) (HPRG).
GN	HRG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RN	SEQUENCE FROM N.A.
RX	MEDLINE=86216149; PubMed=3011081;
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";
RL	Biochemistry 25:2220-2225(1986).
RN	[2]
RN	SEQUENCE FROM N.A.
RA	Wakebayashi S., Takahashi K., Tokunaga F., Koide T.;
RL	Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
RN	[3]
RN	SEQUENCE OF 214-247 FROM N.A.
RX	MEDLINE=94245171; PubMed=8188234;
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,
RA	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";
RL	Genomics 19:195-197(1994).
RN	[4]
RN	SEQUENCE OF 19-27.
RA	TISSUE=Plasma;
RX	MEDLINE=93092937; PubMed=1459097;
RA	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA	Hochstrasser D.F.;
RT	"Plasma protein map: an update by microsequencing.";
RL	Electrophoresis 13:707-714(1992).
CC	-!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS
CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
CC	COAGULATION CASCADE.
CC	-!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-
CC	RESIDUE SEQUENCE (GHPPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.
CC	-!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M13149; AAA52694.1; .  
 DR EMBL; AB005803; BAA21613.1; .  
 DR EMBL; 217218; CAA78925.1; .  
 DR PIR; A01287; AGHUGH.  
 DR SWISS-2DPAGE; P04196; HUMAN.  
 DR MIM; 142640; .  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR SMART; SM00043; CV; 2.  
 KW Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
 FT DOMAIN 19 136 CYSTATIN-LIKE 1.  
 FT DOMAIN 137 254 CYSTATIN-LIKE 2.  
 FT DOMAIN 276 321 PRO-RICH.  
 FT DOMAIN 350 497 PRO/HIS-RICH.  
 FT DISULFID 24 504 BY SIMILARITY.  
 FT DISULFID 78 89 BY SIMILARITY.  
 FT DISULFID 105 126 BY SIMILARITY.  
 FT DISULFID 203 417 BY SIMILARITY.  
 FT DISULFID 218 241 BY SIMILARITY.  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 100.0%; Score 155; DB 1; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-14;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGSVRLPLLRKGEVLPLPEANFPSPPLP 29  
 |||||  
 DB 455 IGSVRLPLLRKGEVLPLPEANFPSPPLP 483

RESULT 2  
 HRG\_BOVIN STANDARD; PRT; 396 AA.  
 AC P3433;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)  
 DE (HPRG) (Fragments).  
 GN HRG.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=93351678; PubMed=8348977;  
 RA Soerensen C.B., Krogh-Pedersen H., Petersen T.E.;  
 RT "Determination of the disulphide bridge arrangement of bovine  
 RT histidine-rich glycoprotein";  
 RL FEBS Lett. 328:285-290(1993).  
 CC -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
 CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
 CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
 CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
 CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
 CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD

CC COAGULATION CASCADE.  
 CC -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
 CC PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-  
 CC RESIDUE SEQUENCE (GHPPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.  
 CC -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.  
 DR PIR; S35687; S35687.  
 KW Glycoprotein; Heparin-binding; Repeat.  
 FT DOMAIN 1 102 CYSTATIN-LIKE 1.  
 FT DOMAIN 103 169 CYSTATIN-LIKE 2.  
 FT DOMAIN 191 238 PRO-RICH.  
 FT DOMAIN 243 368 PRO/HIS-RICH.  
 FT DISULFID 7 375  
 FT DISULFID 56 67  
 FT DISULFID 77 92  
 FT DISULFID 123 297  
 FT DISULFID 137 160  
 FT DISULFID 212 242  
 FT NON\_CONS 52 53  
 FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .).  
 FT NON\_CONS 71 72  
 FT NON\_CONS 78 79  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .).  
 FT NON\_CONS 103 104  
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .).  
 FT NON\_CONS 163 164  
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .).  
 FT NON\_CONS 263 264  
 FT NON\_CONS 303 304  
 FT VARIANT 86 86 S -> R.  
 FT VARIANT 309 309 S -> Q.  
 FT VARIANT 322 322 H -> Y.  
 SQ SEQUENCE 396 AA; 44470 MW; 128A8223499D56FC CRC64;

Query Match 77.4%; Score 120; DB 1; Length 396;  
 Best Local Similarity 78.6%; Pred. No. 2.5e-09;  
 Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 GSVYRLPLLRKGEVLPLPEANFPSPPLP 29  
 |::|||:|||||  
 DB 327 GYIHLRLPSLKGVLPLPEANFPSPSLP 354

RESULT 3  
 HRG\_RABBIT STANDARD; PRT; 526 AA.  
 AC Q28640;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
 DE glycoprotein) (HPRG) (Fragment).  
 GN HRG.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
 RX MEDLINE=96229917; PubMed=8639676;  
 RA Borza D.-B., Tatum F.M., Morgan W.T.;  
 RT "Domain structure and conformation of histidine-proline-rich  
 RT glycoprotein";  
 RL Biochemistry 35:1925-1934(1996).  
 CC -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
 CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
 CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
 CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
 CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
 CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD  
 CC COAGULATION CASCADE.

```

RX MEDLINE=87146477; PubMed=38282835;
RT Sandal N.N., Bojsen K., Marcker K.A.;
RA "A small family of nodule specific genes from soybean.";
RL Nucleic Acids Res. 15:1507-1519(1987).
CC -|- INDUCTION: DURING NODULATION IN LEGUME ROOTS AFTER RHIZOBIUM
CC INFECTON.
CC -|- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensel@isb-sib.ch).
CC -----
CC EMBL; X03979; CAA27618.1; -.
DR EMBL; X05028; CAA28692.1; -.
DR PIR; S09552; S09552.
DR InterPro: IPR003387; Nodulin.
DR Pfam; PF02451; Nodulin; 1.
KW Nodulation.
SQ SEQUENCE 360 AA; 39077 MW; DC4A6D3135EA46DC CRC64;

Query Match          42.9%; Score 66.5; DB 1; Length 360;
Best Local Similarity 40.6%; Pred. No. 0.059;
Matches 13; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY      1 IGSVYRLPPLRKG---EVLPLPEANFSPFLP 29
         :| | | | | :| | | | | | | | |
DB     142 LGSVILPPPIRPGPPLKIIPFGTNIVFPFRP 173

RESULT 5
FUCP_HAEIN STANDARD; PRT; 428 AA.
AC F44776;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE L-fucose permease.
GN FUCP OR H10610.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SE SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Sutton G.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G.J., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -|- FUNCTION: TRANSPORT OF L-FUCOSE INTO THE CELL.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -|- SIMILARITY: BELONGS TO THE FHS (TC 2.1.7) FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



```

DR PROSITE; PS50001; SH2; 1.
KW SH2 domain; Phosphorylation. SH2.
FT DOMAIN 422 530
SQ SEQUENCE 533 AA; 60228 MW; 6DD877782C2E60C3 CRC64;

Query Match 34.5%; Score 53.5; DB 1; Length 533;
Best Local Similarity 56.5%; Pred. No. 5.8;
Matches 13; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 8 PPLR-KGEVLPLPEANFPSPPLP 29
    ||||:| |||| | | | |
DB 389 PPLRSEGRNLPVPNRPQPPSP 411

RESULT 8
MVIN_ECOLI
ID MVIN_ECOLI STANDARD; PRT; 511 AA.
AC P/5932;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Virulence factor mvln homolog.
GN MVIN OR B1069 OR Z1707 OR ECS1447.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT

RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- SURCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MVIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF000208; AAC74153.1; -
CC EMBL; D90743; BAA35877.1; -
CC EMBL; AE005317; AAG55815.1; -
CC EMBL; AF002555; BAB34870.1; -
CC EcoGene; EGI3880; mvln.
CC InterPro; IPR004268; MVIN.
CC Pfam; PF03023; MVIN; 1.
KW Transmembrane; Complete proteome.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 354 374 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT TRANSMEM 481 501 POTENTIAL.
FT SEQUENCE 511 AA; 55267 MW; CB20FE3CFC9419C2 CRC64;
SQ

```

Query Match 32.9%; Score 51; DB 1; Length 511;  
 Best Local Similarity 55.0%; Pred. No. 12;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 VYRLPPLRKGEVLPLPEANF 23  
 ||:| | | | | | | |  
 DB 202 VYOLPHLKKIGMLVLPRI 221

RESULT 9  
 ID TRIA\_PIG STANDARD; PRT; 461 AA.  
 AC P50555;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tumor necrosis factor receptor 1 precursor (p60) (TNF-R1) (TNF-R1)  
 DE (p55).  
 GN TNFRSF1A OR TNFR1.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=96011645; PubMed=7590278;  
 RA Suter B., Pauli U.H.;  
 RT "Cloning of the cDNA encoding the porcine p55 tumor necrosis factor  
 RT receptor.";  
 RL Gene 163:263-266(1995).  
 CC -1- FUNCTION: RECEPTOR FOR TNF-ALPHA. THE ADAPTOR MOLECULE FADD  
 CC RECRUITS CASPASE-8 TO THE ACTIVATED RECEPTOR. THE RESULTING  
 CC AGGREGATE CALLED THE DEATH-INDUCING SIGNALING COMPLEX (DISC)  
 CC PERFORMS CASPASE-8 PROTEOLYTIC ACTIVATION WHICH INITIATES THE  
 CC SUBSEQUENT CASCADE OF CASPASES (ASPARTATE-SPECIFIC CYSTEINE



```
RC STRAIN-LT2 / KK1004;
RA van Slooten J.-C., Okada T., Kutsukake K., Pechere J.-C., Harayama S.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=94259301; PubMed=8200538;
RA Kutsukake K., Okada T., Yokoseki T., Iino T.;
RT "Sequence analysis of the flgA gene and its adjacent region in
RT Salmonella typhimurium, and identification of another flagellar gene,
RT flgN";
RL Gene 143:49-54(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
RC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE MVIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z26133; CA881134.1; -.
DR EMBL; D25292; BAA04980.1; -.
DR EMBL; AE008750; AAL20100.1; -.
DR PIR; S40271; S40271.
DR StyGene; SG10240; mvIN.
DR InterPro; IPR004368; MVIN.
DR Pfam; PF03023; MVIN; 1.
KW Transmembrane; Virulence; Complete proteome.
FT TRANSMEM 44 64
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 396 416 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
FT TRANSMEM 494 514 POTENTIAL.
SQ SEQUENCE 524 AA; 57314 MW; 105BB31AF805E048 CRC64;

Query Match 32.3%; Score 50; DB 1; Length 524;
Best Local Similarity 55.0%; Pred. No. 17;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 VYRLPPLRKGEVLPPEANF 23
DB 215 VIQLPYLKKIGMLVLPRI 234

RESULT 12
PMPH_CHLTR
ID PMPH_CHLTR STANDARD; PRT; 1016 AA.
AC O84880;
DT 16-OCT-2001 (Rel. 40, Created)
```

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmph precursor (Polymorphic membrane
DE protein H).
GN PMPH OR C1872.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001360; AAC68470.1; -.
DR InterPro; IPR003368; DUF145.
DR InterPro; IPR003357; OMP.
DR Pfam; PF02415; DUF145; 1.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1016 PROBABLE OUTER MEMBRANE PROTEIN PMPH.
SQ SEQUENCE 1016 AA; 107904 MW; E691912C3A2BD6F7 CRC64;

Query Match 32.3%; Score 50; DB 1; Length 1016;
Best Local Similarity 69.2%; Pred. No. 36;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 YRLPPLRKGEVLP 17
DB 689 YRLDPQKGEVLP 701

RESULT 13
T3RE_SALTY
ID T3RE_SALTY STANDARD; PRT; 990 AA.
AC P40815;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Type III restriction-modification system StyLT1 enzyme res
DE (EC 3.1.21.5)
GN RES OR STM0358.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT7;
RX MEDLINE=93252265; PubMed=8387444;
RA Dattois V., de Backer O., Colson C.;
RT "Sequence of the Salmonella typhimurium StyLT1 restriction-
RT modification genes: homologues with EcoPI and EcoPI5 type-III R-M
RT systems and presence of helicase domains.";
RL Gene 127:105-110(1993).
```

```

[2]
RN  SEQUENCE FROM N.A.
RP  STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX  MEDLINE=21534948; PubMed=11677609;
RA  McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA  Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA  Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA  Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA  Waterston R., Wilson R.K.;
RT  "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT  LT2";
RL  Nature 413:852-856(2001).
CC  -!- FUNCTION: CLEAVES DNA SOME 25 BASE-PAIRS DOWNSTREAM FROM THE
CC  RECOGNITION SITE. MAY ALSO ACT AS A HELICASE INVOLVED IN
CC  UNWINDING DNA AT THE CLEAVAGE SITE. PROTEIN ONLY REQUIRED FOR
CC  RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME.
CC  -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC  specific double-stranded fragments with terminal 5'-phosphates.
CC  -!- COFACTOR: MAGNESIUM AND ATP.
CC  -!- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD.
CC  -!- SIMILARITY: WITH OTHER TYPE III RES PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL: M90544; -; NOT_ANNOTATED_CDS.
DR  EMBL: AE008711; AAL19312.1; -
DR  PIR: JN0658; JN0658.
DR  REBASE: 1788; STYLT1.
DR  StyGene: SG22222; res.
KW  Hydrolyase; Nuclease; Endonuclease; Restriction system; Helicase;
KW  Complete proteome.
FT  CONFLICT 17 17 A -> R (IN REF. 1).
FT  CONFLICT 293 298 KILRR -> EINPPT (IN REF. 1).
FT  CONFLICT 350 350 S -> T (IN REF. 1).
FT  CONFLICT 420 420 L -> V (IN REF. 1).
FT  CONFLICT 464 465 SG -> R (IN REF. 1).
FT  CONFLICT 536 537 GL -> AV (IN REF. 1).
FT  CONFLICT 945 990 DOVILDAQKFFDMURQINVEFAEATSAVAFVSTINGLI
FT  ESKAN -> IRLFLMNVNSICCVGKISMSLRKDQKAGG
FT  IFYDQWLD (IN REF. 1).
FT  SEQUENCE 990 AA; 113388 MW; 8E821E02A895D16E CRC64;

Query Match 31.9%; Score 49.5; DB 1; Length 990;
Best Local Similarity 44.0%; Pred. No. 41;
Matches 11; Conservative 6; Mismatches 5; Indels 3; Gaps 1;

QY  3. SVYRLPPLRKGEV---LPLPEANFP 24
DB  866 SLXQIPPLRYDSVDPELPLKDYD 890

RESULT 14
HD_FUGRU STANDARD; PRT; 3148 AA.
AC P51112;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Huntingtin (Huntington's disease protein homolog) (HD protein).
GN HD.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=95375788; PubMed=7647794;
RX  Baxendale S., Abdulla S., Elgar G., Buck D., Berks M., Micklem G.,
RA  Durbin R., Bates G., Brenner S., Beck S., Lehrach H.;
RT  "Comparative sequence analysis of the human and pufferfish
RT  Huntingtin's disease genes.";
RL  Nat. Genet. 10:67-76(1995).
CC  -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR
CC  VESICLE FUNCTION.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -!- POLYMORPHISM: THE POLY-GLN REGION (FOUR RESIDUES) DOES NOT APPEAR
CC  TO BE POLYMORPHIC, EXPLAINING THE ABSENCE OF A HD-LIKE DISORDER.
CC  -!- SIMILARITY: CONTAINS 10 HEAT REPEATS.
CC  -!- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL: X82939; CAA58112.1; -
DR  InterPro: IPR000357; HEAT_repeat.
DR  InterPro: IPR000091; Huntingtin.
DR  PRINTS: PR00375; HUNTINGTIN.
DR  PROSITE: PS50077; HEAT_REPEAT; 1.
KW  Repeat.
FT  DOMAIN 148 272 HEAT REPEATS DOMAIN 1.
FT  DOMAIN 701 898 HEAT REPEATS DOMAIN 2.
FT  DOMAIN 1527 1568 HEAT REPEATS DOMAIN 3.
FT  DOMAIN 18 21 POLY-GLN.
FT  DOMAIN 679 682 POLY-ALA.
FT  DOMAIN 1104 1108 POLY-SER.
FT  SEQUENCE 3148 AA; 348932 MW; D9358676B0345243 CRC64;

Query Match 31.9%; Score 49.5; DB 1; Length 3148;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY  1 IGSVYRLPP-LRKGEVLPPEANF 23
DB  1222 LGSFYHLPPYKLYDLKATHANF 1245

RESULT 15
LAT_RAT STANDARD; PRT; 241 AA.
ID LAT_RAT
AC O70601;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Linker for activation of T cells (36 kDa phospho-tyrosine adaptor
DE protein) (pp36) (p36-38).
DE GN LAT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PVG; TISSUE=Natural killer cells;
RX MEDLINE=98197173; PubMed=9529333;
RA Weber J.R., Orstavik S., Torgersen K.M., Danbolt N.C., Berg S.F.,
RA Ryan J.C., Tasken K., Imboden J.B., Vaage J.T.;
RT "Molecular cloning of the cDNA encoding pp36, a tyrosine-
RT phosphorylated adaptor protein selectively expressed by T cells and
RT natural killer cells.";
RL J. Exp. Med. 187:1157-1161(1998).
CC -!- FUNCTION: INVOLVED IN T-CELL ANTIGEN RECEPTOR (TCR) SIGNAL

```



```

CC TRANSDUCTION PATHWAY. MAY PLAY AN IMPORTANT ROLE DOWNSTREAM OF THE
CC ACTIVATION OF PROTEIN TYROSINE KINASES (PTKS) (BY SIMILARITY).
CC -!- SUBUNIT: WHEN PHOSPHORYLATED, INTERACTS DIRECTLY WITH THE P85
CC SUBUNIT OF PHOSPHATIDYL 3-KINASE AND THE SH2 DOMAIN OF GRB2,
CC GRAP, AND PLC-GAMMA-1. ASSOCIATES INDIRECTLY WITH CBL, SOS, VAV,
CC AND SLP-76 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type III membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: NK CELLS.
CC -!- PTM: PHOSPHORYLATED ON TYROSINES BY ZAP-70 AND SYK.
CC PHOSPHORYLATION LEADS TO THE RECRUITMENT OF MULTIPLE SIGNALING
CC MOLECULES. IS ONE OF THE MOST PROMINENTLY TYROSINE-PHOSPHORYLATED
CC PROTEINS DETECTED FOLLOWING TCR ENGAGEMENT (BY SIMILARITY).
CC -!- MISCELLANEOUS: ENGAGEMENT OF KILLER INHIBITORY RECEPTORS (KIR)
CC DISRUPTS THE INTERACTION OF PLC-GAMMA-1 WITH LAT AND BLOCKS TARGET
CC CELL-INDUCED ACTIVATION OF PLC, MAYBE BY INDUCING THE
CC DEPHOSPHORYLATION OF LAT (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AJ001184; CAA04577.1; .
KW Signal-anchor; Transmembrane; Phosphorylation.
FT DOMAIN 1 4 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5 28 SIGNAL-ANCHOR (TYPE III MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 29 241 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 217 224 POLY-GLU.
FT DOMAIN 136 139 PLC-GAMMA-1-BINDING MOTIF.
FT DOMAIN 175 178 GRB2-BINDING MOTIF 1.
FT DOMAIN 195 198 GRB2-BINDING MOTIF 2.
FT MOD_RES 113 113 PHOSPHORYLATION (BY TYR-KINASES)
FT FT (POTENTIAL).
FT MOD_RES 132 132 PHOSPHORYLATION (BY TYR-KINASES)
FT FT (POTENTIAL).
FT MOD_RES 175 175 PHOSPHORYLATION (BY TYR-KINASES)
FT FT (POTENTIAL).
FT MOD_RES 195 195 PHOSPHORYLATION (BY TYR-KINASES)
FT FT (POTENTIAL).
FT MOD_RES 234 234 PHOSPHORYLATION (BY TYR-KINASES)
FT FT (POTENTIAL).
SQ SEQUENCE 241 AA; 26192 MW; FBD5C4D4BBCE4275 CRC64;

```

```

Query Match 31.6%; Score 49; DB 1; Length 241;
Best Local Similarity 42.9%; Pred. No. 10;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 4 VYRLPRLKGEVLPLPEANFP 24
   | ||||: :||: |
Db 70 VTSPPLKROPDLPLPRSPQP 90

```

Search completed: May 24, 2002, 17:03:28  
Job time: 863 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: May 24, 2002, 16:49:05 ; Search time 31.45 Seconds

(without alignments)  
92.336 Million cell updates/sec

Title: US-09-730-379e-1

Perfect score: 424

Sequence: 1 GPRPFCRQIGSVYRLPPLR.....SESCPGKFKSGFPQVSMFEF 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	424	100.0	525	1 HRC_HUMAN	P04196 homo sapien
2	263	62.0	526	1 HRC_RABIT	Q28640 coryctolagus
3	246	58.0	396	1 HRC_BOVIN	P33433 bos taurus
4	80	18.9	428	1 SMA9_MOUSE	O911w5 mus musculu
5	77	18.2	360	1 N044_SOYBN	P04672 glycine max
6	75	17.7	291	1 GDBB_WHEAT	P06659 triticum ae
7	74	17.5	707	1 ZF60_MOUSE	P16374 mus musculu
8	73	17.2	271	1 HMB_XENLA	P14837 xenopus lae
9	72.5	17.1	305	1 H0G1_HORVU	P17990 hordeum vul
10	72	17.0	251	1 GDBO_WHEAT	P08079 triticum ae
11	70.5	16.6	1943	1 PC15_MOUSE	O99fj1 mus musculu
12	69	16.3	494	1 HMC3_DROME	P09081 drosophila
13	68.5	16.2	302	1 GDBX_WHEAT	P21292 triticum ae
14	68	16.0	947	1 M3KE_HUMAN	O99558 homo sapien
15	68	16.0	1355	1 SALM_DROME	P39770 drosophila
16	68	16.0	1745	1 ZOI_MOUSE	P39447 mus musculu
17	67.5	15.9	148	1 GP22_LITCA	O01493 litomosoid
18	67	15.8	241	1 RRRP_HRSVI	P24567 human respi
19	67	15.8	1142	1 ENAM_PIG	O97939 sus scrofa
20	66.5	15.7	481	1 KRL_HSV2H	P13387 herpes simp
21	66.5	15.7	542	1 ZYX_CHICK	Q04584 gallus gall
22	66.5	15.7	1362	1 BRD4_HUMAN	O60808 homo sapien
23	66.5	15.7	1874	1 POLR_KYMYI	P36304 kenedya ye
24	66	15.6	264	1 HOR3_HORVU	P06471 hordeum vul
25	66	15.6	434	1 SMA9_RAT	O54835 rattus norv
26	66	15.6	799	1 F108_YEAST	P40068 saccharomyc
27	66	15.6	971	1 Y029_HUMAN	O15032 homo sapien
28	65.5	15.4	443	1 FB14_CRIGR	O55058 cricoetulus
29	65.5	15.4	533	1 YD3D_SCHPO	O10277 schizosacch
30	65.5	15.4	1528	1 KEM1_YEAST	P22447 saccharomyc
31	65.5	15.4	1607	1 MIPR_LYMAST	Q23410 lymaea sta
32	65	15.3	411	1 PAL1_HUMAN	O9um63 homo sapien
33	65	15.3	465	1 SMA5_HUMAN	O99717 homo sapien

34	65	15.3	465	1 SMA5_MOUSE	P97454 mus musculu
35	65	15.3	465	1 SMA5_RAT	O911v3 rattus norv
36	65	15.3	782	1 SMA8_MOUSE	O62179 mus musculu
37	64	15.1	293	1 HOR1_HORVU	P06470 hordeum vul
38	64	15.1	490	1 CLK3_HUMAN	P49761 homo sapien
39	64	15.1	490	1 CLK3_MOUSE	O35492 mus musculu
40	64	15.1	490	1 CLK3_RAT	O63117 rattus norv
41	63.5	15.0	307	1 GDA9_WHEAT	P18573 triticum ae
42	63.5	15.0	443	1 FB14_HUMAN	O95967 homo sapien
43	63.5	15.0	633	1 LA17_YEAST	O12446 saccharomyc
44	63	14.9	260	1 HXC9_HUMAN	P31274 homo sapien
45	63	14.9	260	1 HXC9_MOUSE	P09633 mus musculu

## ALIGNMENTS

RESULT 1

HRG\_HUMAN STANDARD; PRT; 525 AA.

AC P04196:

DT 20-MAR-1987 (rel. 04, Created)

DT 20-MAR-1987 (rel. 04, Last sequence update)

DT 16-OCT-2001 (rel. 40, Last annotation update)

DE Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).

GN HRG.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

XX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86216149; PubMed=3011081;

RA Koide T., Foster D.C., Yoshitake S., Davie E.W.;

RT "Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";

RL Biochemistry 25:2220-2225(1986).

RN [2]

RP SEQUENCE FROM N.A.

RA Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 214-247 FROM N.A.

RX MEDLINE=94245171; PubMed=8188234;

RA Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W., Biondini L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;

RT "Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";

RL Genomics 19:195-197(1994).

RN [4]

RP SEQUENCE OF 19-27.

RC TISSUE=Plasma;

RX MEDLINE=93092937; PubMed=1459097;

RA Hughes G.J., Fruhlinger S., Paquet N., Ravier F., Pasquali C., Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B., Hochstrasser D.F.;

RT "Plasma protein map: an update by microsequencing.";

RL Electrophoresis 13:707-714(1992).

CC -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN, AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN MAY MEDATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD COAGULATION CASCADE.

CC -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.

CC -!- SIMILARITY: CONTAINS 2 CYSTEINE-LIKE DOMAINS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 -----  
 DR EMBL: M3149; AA052694.1; -;  
 DR EMBL: AB005803; BAA21613.1; -;  
 DR EMBL: 217218; CAA78925.1; -;  
 DR PIR: A01287; KGHUGH.  
 DR SWISS-2DPAGE: P04196; HUMAN.  
 DR MIM: 142640; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; cystatin; 1.  
 DR SMART: SM00043; CY; 2.  
 KM Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 525  
 FT DOMAIN 19 136  
 FT DOMAIN 137 254  
 FT DOMAIN 276 321  
 FT DOMAIN 350 497  
 FT DISULFID 24 504  
 FT DISULFID 78 89  
 FT DISULFID 105 126  
 FT DISULFID 203 417  
 FT DISULFID 218 241  
 FT CARBOHYD 63 63  
 FT CARBOHYD 87 87  
 FT CARBOHYD 125 125  
 FT CARBOHYD 344 344  
 FT CARBOHYD 345 345  
 SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;  
 -----  
 Query Match 100.0%; Score 424; DB 1; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-37;  
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 1 GPRFHCROIGSVYRLPLRKGEVLRPLEANFSPFRLHHKHNPLKPDNPFPQSVSSSCP 60  
 DB 446 GPRFHCROIGSVYRLPLRKGEVLRPLEANFSPFRLHHKHNPLKPDNPFPQSVSSSCP 505  
 QY 61 GKFKSGFPQVSMFPT 75  
 DB 506 GKFKSGFPQVSMFPT 520  
 -----  
 RESULT 2  
 HRG\_RABIT STANDARD; PRT; 526 AA.  
 AC Q28640;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
 DE glycoprotein) (HPRG) (Fragment).  
 GN HRG.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID:9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
 RC TISSUE-Serum;  
 RA MEDLINE:96229917; PubMed:8639676;  
 RA Borza D., B., Tatum F.M., Morgan W.M.;  
 RT "Domain structure and conformation of histidine-proline-rich  
 RT glycoprotein."  
 RL Biochemistry 35:1925-1934 (1996).  
 CC 1- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
 CC HEME, DIES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE

CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
 CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
 CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
 CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD  
 CC COAGULATION CASCADE.  
 CC 1- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
 CC RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH  
 CC REGION.  
 CC 1- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 -----  
 DR EMBL: U32189; AAC48516.1; -;  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; cystatin; 1.  
 DR SMART: SM00043; CY; 2.  
 KM Glycoprotein; Heparin-binding; Repeat; Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL 1 8  
 FT CHAIN 1 526  
 FT DOMAIN 9 126  
 FT DOMAIN 127 243  
 FT DOMAIN 251 296  
 FT DOMAIN 329 498  
 FT DISULFID 14 505  
 FT DISULFID 68 79  
 FT DISULFID 95 116  
 FT DISULFID 193 415  
 FT DISULFID 207 230  
 FT DISULFID 272 302  
 FT CARBOHYD 115 115  
 FT CARBOHYD 192 192  
 FT CARBOHYD 240 240  
 FT CARBOHYD 310 310  
 FT CARBOHYD 485 485  
 FT SITE 303 304  
 FT SITE 421 422  
 SQ SEQUENCE 526 AA; 58877 MW; 810F23D367D93D42 CRC64;  
 -----  
 Query Match 62.0%; Score 263; DB 1; Length 526;  
 Best Local Similarity 66.2%; Pred. No. 2.1e-20;  
 Matches 49; Conservative 7; Mismatches 18; Indels 0; Gaps 0;  
 -----  
 QY 1 GPRFHCROIGSVYRLPLRKGEVLRPLEANFSPFRLHHKHNPLKPDNPFPQSVSSSCP 60  
 DB 447 GPRFHCROIGSVYRLPLRKGEVLRPLEANFSPFRLHHKHNPLKPDNPFPQSVSSSCP 506  
 QY 61 GKFKSGFPQVSMFPT 74  
 DB 507 EEFNGEFAQLSKFF 520  
 -----  
 RESULT 3  
 HRG\_BOVIN STANDARD; PRT; 396 AA.  
 AC P33433;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)  
 DE (HPRG) (Fragments).  
 GN HRG.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=93351678; PubMed=8348977;  
 RA Soerensen C.B., Krogh-Pedersen H., Petersen T.E.;  
 RT "Determination of the disulphide bridge arrangement of bovine  
 histidine-rich glycoprotein.";  
 RL FEBS Lett. 328:285-290(1993).  
 CC -1- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
 HEHE, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
 FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
 AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS  
 CC HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
 CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD  
 CC COAGULATION CASCADE.  
 CC -1- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
 CC PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-  
 CC RESIDUE SEQUENCE (GHHH, CONSENSUS) FORM A HISTIDINE-RICH REGION.  
 CC -1- SIMILARITY: CONTAINS 2 CYPSTATIN-LIKE DOMAINS.  
 DR PIR: S35687; S35687.  
 KM Pfam: PF00031; cystatin; 1.  
 KW Glycoprotein; Heparin-binding; Repeat.  
 FT DOMAIN 1 102 CYPSTATIN-LIKE 1.  
 FT DOMAIN 103 169 CYPSTATIN-LIKE 2.  
 FT DOMAIN 191 238 PRO-RICH.  
 FT DOMAIN 243 368 PRO/HIS-RICH.  
 FT DISULFID 7 375  
 FT DISULFID 56 67  
 FT DISULFID 77 92  
 FT DISULFID 123 297  
 FT DISULFID 137 160  
 FT DISULFID 212 242  
 FT NON\_CONS 52 53  
 FT CARBOHYD 70 70  
 FT NON\_CONS 71 72  
 FT NON\_CONS 78 79  
 FT CARBOHYD 91 91  
 FT NON\_CONS 103 104  
 FT CARBOHYD 122 122  
 FT NON\_CONS 163 164  
 FT CARBOHYD 220 220  
 FT NON\_CONS 263 264  
 FT NON\_CONS 303 304  
 FT VARIANT 86 86  
 FT VARIANT 309 309  
 FT VARIANT 322 322  
 FT VARIANT H -> Y.  
 SQ SEQUENCE 396 AA; 44470 MW; 128A8223499DE6FC CRC64;  
 Query Match 58.0%; Score 246; DB 1; Length 396;  
 Best Local Similarity 64.3%; Pred. No. 9,3e-19;  
 Matches 45; Conservative 8; Mismatches 17; Indels 0; Gaps 0;  
 QY 5 FHCRQIGSVYRLPPLPEANFPSPFLPHKHPDLKPDNPFPQSPVSESCPGKFK 64  
 DB 321 FHMRRPGYHRLPSLKKSGVLPLEANFPSPFLPHKHPDLKPDNPFPQSPVSESCPGKFK 380  
 QY 65 SGFPQVSMFP 74  
 DB 381 IKFLHISKFF 390  
 RESULT 4  
 ID SMA9\_MOUSE STANDARD; PRT; 428 AA.  
 AC Q9J1M5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mothers against decapentaplegic homolog 9 (SMAD9) (Mothers against  
 DE DMP homolog 9) (Smad9) (Smad8).  
 GN MADH9 OR MADH8 OR SMAD8.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=20275476; PubMed=10814522;  
 RA Kawai S., Faucheu C., Galilea S., Spinnella-Jeagle S., Atti A.,  
 RA Baron R., Roman-Roman S.;  
 RT "Mouse Smad8 phosphorylation downstream of BMP receptors ALK-2, ALK-3,  
 RT and ALK-6 induces its association with Smad4 and transcriptional  
 RT activity.";  
 RL Biochem. Biophys. Res. Commun. 271:682-687(2000).  
 CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR ACTIVATED BY BMP (BONE  
 CC MORPHOGENETIC PROTEINS) TYPE I RECEPTOR KINASE. SMAD8 IS A  
 CC RECEPTOR-REGULATED SMAD (R-SMAD). HAS BEEN SHOWN TO BE ACTIVATED  
 CC BY ACTIVIN TYPE I RECEPTOR-LIKE KINASES (ALK-2, ALK-3, ALK-6)  
 CC WHICH STIMULATE HETEROMERIZATION BETWEEN SMAD8 AND SMAD4. MAY PLAY  
 CC A ROLE IN OSTEOBLAST DIFFERENTIATION AND MATURATION.  
 CC -1- SUBUNIT: INTERACTION WITH THE CO-SMAD SMAD4. INTERACTS WITH  
 CC PEBP2-ALPHA SUBUNIT (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: IN THE CYTOPLASM IN THE ABSENCE OF LIGAND;  
 CC MIGRATION TO THE NUCLEUS WHEN COMPLEXED WITH THE CO-SMAD SMAD4.  
 CC -1- PTM: PHOSPHORYLATED ON SERINE BY BMP (BONE MORPHOGENETIC PROTEINS)  
 CC TYPE I RECEPTOR KINASE AND ACTIVIN TYPE I RECEPTOR-LIKE KINASES  
 CC (ALK-2, ALK-3 AND ALK-6).  
 CC -1- SIMILARITY: BELONGS TO THE DWARFIN/SMAD FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DWA/MH2 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF175408; AAF77079.1; -.  
 DR MGD: MGI:1859993; Madh9.  
 DR InterPro: IPR001132; DwarfIn.  
 DR InterPro: IPR003619; DwarfIn\_A.  
 DR Pfam: PF00968; DwarfIn; 1.  
 DR SMART: SM00523; DwarfIn; 1.  
 DR SMART: SM00524; DMB; 1.  
 KW SMART; SM00524; DMB; 1.  
 KW Transcription regulation; Multigene family; Phosphorylation.  
 FT DOMAIN 29 138 DWA.  
 FT DOMAIN 232 404 DMB.  
 FT DOMAIN 43 49 POLY-LYS.  
 FT MOD\_RES 425 425 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 426 426 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 428 428 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 428 AA; 48314 MW; 5BEC95CCBAE7B40F CRC64;  
 Query Match 18.9%; Score 80; DB 1; Length 428;  
 Best Local Similarity 37.3%; Pred. No. 0.26;  
 Matches 28; Conservative 7; Mismatches 28; Indels 12; Gaps 6;  
 QY 4 PFHCRQIGSVYRLPPLPEANFPSPFLPHKHPDLKPDNPFPQSPVSESCPGKFK 58  
 DB 128 PYHYRRVGVTV-LPVLVYRHSBYNDQLSLAKFNSASL--HSEPLMPNATYPPSPFQSS 184  
 QY 59 -CPGKFKSG--FPQ 69  
 DB 185 LCPAPPSPGHVFPQ 199  
 RESULT 5  
 ID NO44\_SOYBN STANDARD; PRT; 360 AA.  
 AC P04672; P07127;  
 DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Nodulin 44 (N-44) (Nodulin E27).  
 OS Glycine max (Soybean).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Psoraleae; Glycine.  
 OC NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root nodules;  
 RA Sengupta-Gopalan C., Pitas J.W., Thompson D.V., Hoffman L.M.;  
 RT "Expression of host genes during root nodule development in  
 RL soybeans.";  
 RL Mol. Gen. Genet. 203:410-420(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=87146477; PubMed=3822835;  
 RA Sandali N.N., Bojsen K., Marcker K.A.;  
 RT "A small family of nodule specific genes from soybean."  
 RL Nucleic Acids Res. 15:1507-1519(1987).  
 CC -I- INDUCTION: DURING NODULATION IN LEGUME ROOTS AFTER RHIZOBIUM  
 CC INFECTION.  
 CC -I- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.  
 -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; X03979; CAA27618.1; -;  
 DR EMBL; X05028; CAA28692.1; -;  
 DR PIR; S09552; S09552.  
 DR InterPro; IPR003387; Nodulin.  
 DR Pfam; PF02451; Nodulin; 1.  
 DR Nodulation.  
 KW SEQUENCE 360 AA; 39077 MW; DC4A6D3135EA46CD CRC64;  
 SO

Query Match 18.2%; Score 77; DB 1; Length 360;  
 Best Local Similarity 36.2%; Pred. No. 0.44;  
 Matches 17; Conservative 7; Mismatches 13; Indels 10; Gaps 2;  
 Y 10 IGSYVRLPRLRG--EVLPLPEANFSPRLPHKHNLKPNQNPQQ 53  
 D 142 LGSYVRLPRLRG--EVLPLPEANFSPRLPHKHNLKPNQNPQQ 181

RESULT 6  
 GDBB\_WHEAT  
 ID GDBB\_WHEAT STANDARD; PRT; 291 AA.  
 AC P06639;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Gamma-gliadin B precursor.  
 OS Triticum aestivum (Wheat).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 CC Triticeae; Triticum.  
 OC NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=86301876; PubMed=3017812;  
 RA Rafalski J.A.;  
 RT "Structure of wheat gamma-gliadin genes."  
 RL Gene 43:221-229(1986).  
 CC -I- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.  
 CC -I- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY  
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO

```
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR  
CC NEAR DIRECT REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce,  
CC or send an email to licenses@isb-sib.ch).  
-----  
DR EMBL_M13713; AAA34274.1; -;  
DR PIR_A25632; EEMTG.  
DR InterPro_IPR003612; AAI.  
DR InterPro_IPR001768; Cereals_ryp.amyl_inh.  
DR InterPro_IPR001954; Glia_gluclnln.  
DR Pfam_Pf00234; tryp.alpha_amy1.1.  
DR PRINTS_PR00208; GLIADGLOTEM.  
DR SMART_SM00499; AAI; 1.  
KW Seed storage protein; Repeat; signal; Multigene family.  
FT SIGNAL 1 19  
FT CHAIN 20 291 GAMMA-GLIADIN B.  
SQ SEQUENCE 291 AA; 32967 MW; 85732C2A2EFC041 CRC64;  
  
Query Match 17.7%; Score 75; DB 1; Length 291;  
Best Local Similarity 30.2%; Pred.No. 0.57;  
Matches 16; Conservative 7; Mismatches 24; Indels 6; Gaps 1  
  
OY 23 EVLPLEPEANFESFLPHHKHKLKPD-----NQPPQSVSESCPGKFSGFPQ 69  
Db 50 QIFPQPQQTFEPHQPOQOFEPQOPQOOQLDPLRPQPFQDPQGPYPQOPQOPFPQ 102  
::|::||| | : ||| | : : ||| |  
ZEF6_MOUSE STANDARD: PRT; 707 AA.  
ID ZF60_MOUSE  
AC PI6374; O61135;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Zinc finger protein 60 (Zfp-60) (Zinc finger protein Mfg-3).  
OS Zfp60 OR MFG3.  
GN Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxId=10090;  
PI [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=96244506; PubMed=8674531;  
RX Perez M., Rompatto G., Corbi N., de Gregorio L., Draganti T.A.,  
RA Passanantti C.;  
RT "zfzpf60, a mouse zinc finger gene expressed transiently during in  
RT vitro muscle differentiation." ;  
RL FEBS Lett. 387:117-121(1996).  
LN [2]  
RN SEQUENCE OF 325-629 FROM N.A.  
RP STRAIN=CD-1; Tissue=Skeletal muscle;  
RC MEDLINE=90083278; PubMed=2512579;  
RX Passanantti C., Felsenst A., Carnuso M., Amati P.;  
RA "Mouse genes coding for 'zinc-finger'-containing proteins:  
RT characterization and expression in differentiated cells.";  
RT Proc. Natl. Acad. Sci. U.S.A. 86:9417-9421(1989).  
CC -1- FUNCTION: MAY HAVE A ROLE DURING DIFFERENTIATION PROCESSES.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- TISSUE SPECIFICITY: EXPRESSED WIDELY AND EVENLY IN MOST ADULT  
CC MOUSE TISSUES.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS POSITIVELY REGULATED UPON  
CC DIFFERENTIATION, AND IS NOT RELATED TO THE CELL CYCLE.  
CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF CZHR-TYPE ZINC-FINGER PROTEINS.  
CC -1- SIMILARITY: CONTAINS 1 KRB DOMAIN.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U48721; AAB06876.1; -  
 DR EMBL: M28515; AAA39533.1; -  
 DR PIR: C39240; C39240.  
 DR HSSP: P07248; 1PAA.  
 DR MGD: MGI:99207; 2F60.  
 DR InterPro: IPR00109; KRAB.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF01352; KRAB.1.  
 DR Pfam: PF00096; Zf-C2H2; 19.  
 DR PRINTS: PR00048; ZINCFINGER.  
 DR SMART: SM00349; KRAB.1.  
 DR SMART: SM00355; Znf-C2H2; 19.  
 DR PROSITE: PS00805; KRAB.1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 18.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 19.  
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein.  
 FT DOMAIN 14 86  
 FT ZN\_FING 173 702 ZINC\_FINGERS.  
 FT ZN\_FING 201 223 C2H2-TYPE.  
 FT ZN\_FING 229 251 C2H2-TYPE.  
 FT ZN\_FING 288 310 C2H2-TYPE.  
 FT ZN\_FING 316 338 C2H2-TYPE.  
 FT ZN\_FING 344 366 C2H2-TYPE.  
 FT ZN\_FING 372 394 C2H2-TYPE.  
 FT ZN\_FING 400 422 C2H2-TYPE.  
 FT ZN\_FING 428 450 C2H2-TYPE.  
 FT ZN\_FING 456 478 C2H2-TYPE.  
 FT ZN\_FING 484 506 C2H2-TYPE.  
 FT ZN\_FING 512 534 C2H2-TYPE.  
 FT ZN\_FING 540 562 C2H2-TYPE.  
 FT ZN\_FING 568 590 C2H2-TYPE.  
 FT ZN\_FING 596 618 C2H2-TYPE.  
 FT ZN\_FING 624 646 C2H2-TYPE.  
 FT ZN\_FING 652 674 C2H2-TYPE.  
 FT ZN\_FING 680 702 C2H2-TYPE.  
 SO SEQUENCE 707 AA; 82273 MW; 7E0E404C4F0911DE CRC64;

Query Match 17.5%; Score 74; DB 1; Length 707;  
 Best Local Similarity 31.8%; Pred. No. 1.9;  
 Matches 27; Conservative 12; Mismatches 30; Indels 16; Gaps 7;

QY 1 GRPFRGCRIGSVYRLPLP-----RKGEVLP-EPANF--PSFPLP-----HHKHPKLPDNO 49  
 DB 197 GRPFRGCRIGSVYRLPLP-----RKGEVLP-EPANF--PSFPLP-----HHKHPKLPDNO 255  
 QY 50 PFGSVSESCPGKFKSGFPOVSMEF 74  
 DB 256 PFE---CECGKSKFS-FNRISTLF 276

## RESULT 8

ID HM8\_XENLA STANDARD; PRT; 271 AA.  
 AC P14837;  
 DT 01-APR-1990 (rel. 14, Created)  
 DT 01-NOV-1995 (rel. 32, Last sequence update)  
 DT 01-NOV-1997 (rel. 35, Last annotation update)  
 DE Homeobox protein 8 (XLB08-8).  
 GN XLB08.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=83355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gama L.;  
 RN Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE OF 165-271 FROM N.A.  
 RX MEDLINE=90091818; PubMed=2574662;  
 RT Wright C.V.E., Schmeissberg P., de Roberts E.M.;  
 RT "XLB08: a novel Xenopus homeo protein restricted to a narrow band  
 RT of endoderm."  
 RL Development 105:787-794(1989).  
 CC -1 SUBCELLULAR LOCATION: Nuclear.  
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN THE ANTEROPOSTERIOR BAND OF THE  
 CC ENDODERM.  
 CC -1 SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.  
 CC IPI1/XLB08 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X16849; CAA34746.1; -  
 DR PIR: S07818; S07818.  
 DR HSSP: P02833; 9ANT.  
 DR InterPro: IPR001827; Antennapedia.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00025; ANTENNAPEDIA.  
 DR PRINTS: PR00024; HOMEOBOX.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEOBOX\_1; 1.  
 DR PROSITE: PS00071; HOMEOBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT DOMAIN 77 97  
 FT ZN\_FING 119 124 POLY-HIS.  
 FT ZN\_FING 147 206 ANTP-TYPE HEXAPEPTIDE.  
 FT CONFLICT 256 256 HOMEBOX.  
 FT CONFLICT 256 256 P -> S (IN REF. 2).  
 SO SEQUENCE 271 AA; 31406 MW; 024C41F509001037 CRC64;

Query Match 17.2%; Score 73; DB 1; Length 271;  
 Best Local Similarity 26.0%; Pred. No. 0.86;  
 Matches 20; Conservative 13; Mismatches 22; Indels 22; Gaps 4;

QY 12 SVYRLPLRKGEVLP-EPANF-----PSFPLPHKHPKLPDNO-----PPF 52  
 DB 62 SVYRLPLRKGEVLP-EPANF-----PSFPLPHKHPKLPDNO-----PPF 121  
 QY 53 ---QSVSESCPGKFKSG 66  
 DB 122 WKKSTKSHTKGQWTDG 138

## RESULT 9

ID HOG1\_HORVU STANDARD; PRT; 305 AA.  
 AC P17930;  
 DT 01-NOV-1990 (rel. 16, Created)  
 DT 01-NOV-1990 (rel. 16, Last sequence update)  
 DT 01-OCT-1996 (rel. 34, Last annotation update)  
 DE Gamma-hordein 1 precursor.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]

**RP** SEQUENCE FROM N.A.  
**RC** STRAIN-CV CARINA;  
RA Cameron-Mills V., Brandt A.;  
RL "A gamma-hordein gene";  
RM Plant Mol. Biol. 11;449-461(1988).  
CC  
CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC (AS GLOBULES) AND VACUOLAR  
CC (AS PROTEIN BODIES).  
CC  
CC -1 TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.  
CC  
CC -1 DOMAIN: SULFUR RICH HORDEIN WHICH POSSES AN N-TERMINAL HALF  
CC COMPOSED OF PROLINE-GUTAMINE BLOCKS ORGANIZED IN REPEATING UNITS  
CC AND A C-TERMINAL HALF WHERE THE REPEATS ARE DISPERSED AND LESS  
CC CONSERVED.

-1 SIMILARITY: TO GAMMA-GLIADIN FROM WHEAT, AND A GAMMA-SECALIN FROM  
RYE, AND LESS TO A BI HORDEIN FROM BARLEY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement ([see http://www.isb.ch/announce/](http://www.isb.ch/announce/), or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).-----

**DR** EMBL; X13508; CAAB31861.1; -;  
**DR** EMBL; M36578; AAA32955.1; -;  
**DR** PIR; S08312; S08312.  
**DR** InterPro; IPR003612; AA1.  
**DR** InterPro; IPR001768; Cereal\_tryp\_amyl\_Inv.  
**DR** InterPro; IPR001954; Glia\_glutennn.  
**DR** InterPro; IPR000528; Plant\_LTP.  
**DR** Pfam; PF00279; LTP; 1.  
**DR** Pfam; PF00234; tryp\_alpha-amyl; 1.  
**DR** PRINTS; PR00208; GLYADGJUTEN.  
**DR** SMART; SM00499; AAI; 1.  
**KW** Seed storage protein; Signal; Multigene family.  
**FT** SIGNAL 1..19  
**FT** CHAIN 1..305 GAMMA-HORDEIN 1.  
**SQ** SEQUENCE 305 AA; 34737 MW; 60B038533EFB24AD CRC64;

Query Match 17.1%; Score 72.5; DB 1; Length 305;  
Best Local Similarity 35.7%; Pred. No.1.1'  
Matches 20; Conservative 6; Mismatches 21; Indels 9; Gaps 3.

**Dy** 26 PLP-EANFP-----SPFLPHHKHLPKDPNQPFPPSVSSSCPGKRSGFGPVYSMPF 74  
| | : | | : | : | ||||| : : : ||| : |  
**Dd** 36 PYPSQQPFISQSQQQPPOQPPFOOP-QGFPPQS-OQQCLIQAPQHNPDPPOOF 89

**RESULT 10**

ID	GDBO_WHEAT	STANDARD:	PRT:	251 AA.
AC	P08079.			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	01-NOV-1990 (Rel. 16, Last annotation update)			
DE	Gamma-gliadin precursor (Fragment).			
OS	Triticum aestivum (wheat).			
OC	Euarystida, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophytas; Magnoliophyta; Lilopsids; Poales; Poaceae; Poideae;			
OC	Triticeae; Triticum.			
OX	NCBI_TaxId=4565;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Schneets K., Kataiski J.A., Hedgcock C., Soell D.G.;			
RT	"Heptaepptide repeat structure of a wheat gamma-gliadin.";			
RL	Plant Sci. Lett. 37:221-225(1985).			
CC	-1 FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.			
CC	-1 MISCELLANEOUS: THE GAMMA-GLINDINS CAN BE DIVIDED INTO 3 HOMOLOGY CLASSES. SEQUENCE DIVERGENGE BETWEEN THE CLASSES IS DUE TO SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR DIRECT REPEATS.			
CC				

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch). CC or send an email to license@sib-sib.ch).
```

---

```
DR EMBL; M16060; AAA34288.1; -.
DR PIR; PS0094; PS0094.
DR InterPro; IPR001768; Cereal_1tryp_aml1_inh.
DR Pfam; PF00234; tryp_alpha_aml1; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 19
FT CHAIN 20 >251 GAMMA-GLIADIN.
FT NON_TER 251
FT SEQUENCE 251 AA; 29054 MW; 5A3F12C36C825EAD CRC64;
```

---

```
Query Match          17.0%; Score 72; DB 1; Length 251;
Best Local Similarity 30.2%; Pred. No. 1;
Matches 16; Conservative 6; Mismatches 25; Indels 6; Gaps 1.
```

---

```
OY      23 EVLPPEANPFSPFLPHHKHPLKPD-----NQPPQSYSESCPKFKFSGFQQ 69
           : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       50 QTFPQQQTFFHQDQDQFPQPQDQLRQDPQFPQPQDQPYQPQDQDQFPQO 102
```

---

```
RESULT 11
PC15_MOUSE STANDARD: PRT; 1943 AA.
AC Q99PJ1;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protocadherin 15 precursor.
OS PCDH15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=20578757; PubMed=11138007;
RA Alagremam K.N., Murcia C.L., Kwon H.Y., Pawlowski K.S., Wright C.G.,
RA Wojczyk R.P.;
RT "The mouse Ames waltzer hearing-loss mutant is caused by mutation of
RL Pcdh15, a novel protocadherin gene.";
RN Nat. Genet. 27:99-102(2001).
RN [2]
RX TISSUE SPECIFICITY.
RX PubMed=11429292;
RA Murcia C.L., Wojczyk R.P.;
RT "Expression of Pcdh15 in the inner ear, nervous system and various
RL epithelia of the developing embryo.";
RN Mech. Dev. 105:163-166(2001).
CC - FUNCTION: Calcium-dependent cell-adhesion protein. Essential for
CC maintenance of normal retinal and cochlear function. Required for
CC inner ear neuroepithelial cell elaboration.
CC - SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC - TISSUE SPECIFICITY: Expressed in brain and sensory epithelium of
CC the developing inner ear. Also found in the spleen, developing
CC eye, dorsal root ganglion, dorsal aspect of neural tube, floor
CC plate and ependymal cells adjacent to the neural canal.
CC - DEVELOPMENTAL STAGE: Highest level of expression is detected at
CC embryonic day 16.
CC - DISEASE: Defects in PCDH15 are the cause of the Ames waltzer (av)
CC phenotype. It is characterized by deafness and a balance disorder,
CC associated with the degeneration of inner ear neuroepithelia.
CC - SIMILARITY: CONTAINS 11 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF281899; AAG53891.1; -  
 DR HSSP: P09803; ISUH.  
 DR MGI: 1891428; Pcdh15.  
 DR InterPro: IPR002126; Caderlin.  
 DR Pfam: PF00028; caderlin; 10.  
 DR PRINTS: SM00112; CA; 11.  
 DR SMART: PS00232; CADERLIN\_1; 4.  
 DR PROSITE: PS00268; CADERLIN\_2; 11.  
 DR PROSITE: PS00268; CADERLIN\_2; 11.  
 KM Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal; Multigene Family; Vslon.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1943  
 FT DOMAIN 27 1381  
 FT TRANSMEM 1382 1402  
 FT DOMAIN 1403 1943  
 FT DOMAIN 45 152  
 FT DOMAIN 153 270  
 FT DOMAIN 283 400  
 FT DOMAIN 401 514  
 FT DOMAIN 515 621  
 FT DOMAIN 622 722  
 FT DOMAIN 724 824  
 FT DOMAIN 825 931  
 FT DOMAIN 932 1040  
 FT DOMAIN 1042 1149  
 FT DOMAIN 1150 1264  
 FT DOMAIN 1437 1448  
 FT DOMAIN 1772 1778  
 FT DOMAIN 1804 1812  
 FT CARBOHYD 57 57  
 FT CARBOHYD 102 102  
 FT CARBOHYD 206 206  
 FT CARBOHYD 424 424  
 FT CARBOHYD 424 424  
 FT CARBOHYD 564 564  
 FT CARBOHYD 667 667  
 FT CARBOHYD 729 729  
 FT CARBOHYD 773 773  
 FT CARBOHYD 826 826  
 FT CARBOHYD 856 856  
 FT CARBOHYD 1069 1069  
 FT CARBOHYD 1089 1089  
 FT CARBOHYD 1180 1180  
 SQ SEQUENCE 1943 AA; 214816 MW; E3D7DB9F5F738652 CRC64;

Query Match 16.6%; Score 70.5; DB 1; Length 1943;  
 Best Local Similarity 43.2%; Pred. No. 13;  
 Matches 19; Conservative 2; Mismatches 12; Indels 11; Gaps 2;

QY 17 PRLRKGEVLPLRANFSPRLPHKHKLRDNPPOSVSESCP 60  
 DB 1754 P-----LRRPPIAFTTFLP-----PLSPNPPLPQLVTFSLP 1786

RESULT 12  
 HMBC\_DROME STANDARD; PRT; 494 AA.  
 AC P09061;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeotic bicoid protein (PRD-4).  
 GN BCD.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=OREGON-R; TISSUE=Embryo;  
 RX MEDLINE=89005064; PubMed=2901954;  
 RA Berleth T., Burri M., Thoma G., Bopp D., Richstein S., Frigerio G.,  
 RA Noll M., Nusslein-Volhard C.;  
 RT "The role of localization of bicoid RNA in organizing the anterior  
 RT pattern of the Drosophila embryo."  
 RL EMO J. 7:1749-1756(1988).  
 RN [2]  
 RP SEQUENCE OF 86-156 FROM N.A.  
 RX MEDLINE=87051745; PubMed=2877746;  
 RA Frigerio G., Burri M., Bopp D., Baumgartner S., Noll M.;  
 RT "Structure of the segmentation gene paired and the Drosophila PRD  
 RT gene set as part of a gene network."  
 RL Cell 47:735-746(1986).  
 RN [3]  
 RP POSSIBLE RNA-BINDING DOMAIN.  
 RX MEDLINE=89324068; PubMed=2752425;  
 RA Rebagliati M.;  
 RT "An RNA recognition motif in the bicoid protein."  
 RL Cell 58:231-232(1989).  
 CC -1- FUNCTION: BICOID IS A POLARITY PROTEIN THAT PROVIDES POSITIONAL  
 CC CUES FOR THE DEVELOPMENT OF HEAD AND THORACIC SEGMENTS. BCD  
 CC REGULATES THE EXPRESSION OF ZYGOTIC GENES, POSSIBLY THROUGH ITS  
 CC HOMODOMAIN, AND INHIBITS THE ACTIVITY OF OTHER MATERNAL GENE  
 CC PRODUCTS. IT IS POSSIBLE THAT BCD ALSO BINDS RNA.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: MATERNAL EXPRESSION IS AN ANTERIOR CAP  
 CC CONCENTRATED IN THE CORTICAL CYTOPLASM.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.  
 CC "BICOID" SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 "PRD MOTIF".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X14458; CAA32627.1; -  
 CC EMBL: X14459; CAB37631.1; -  
 CC EMBL: M14549; AAA28385.1; -  
 CC EMBL: X14460; CAA32629.1; -  
 CC EMBL: X07870; CAA30720.1; -  
 CC EMBL: K03517; AAA28391.1; -  
 CC PIR: S00835; WJFBC.  
 CC HSSP: P22808; 1VND.  
 CC TRANSFAC: T00063; -  
 CC flybase: Fggn0000166; bcd.  
 CC InterPro: IPR001356; Homeobox.  
 CC Pfam: PF00046; homeobox; 1.  
 CC SMART: SM00389; HOX; 1.  
 CC PROSITE: PS00027; HOMEOBOX\_1; 1.  
 CC PROSITE: PS50071; HOMEOBOX\_2; 1.  
 KM Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation; RNA-binding; Alternative splicing.  
 FT DOMAIN 12 40  
 FT DNA\_BIND 97 156  
 FT DOMAIN 260 294  
 FT DOMAIN 433 440  
 FT VARSPLIC 81 85  
 FT VARSPLIC 56 400  
 FT CONFLICT 298 298  
 SQ SEQUENCE 494 AA; 54493 MW; 561D84A2CE711FD3 CRC64;

Query Match 16.3%; Score 69; DB 1; Length 494;  
 Best Local Similarity 50.0%; Pred. No. 4.3;  
 Matches 15; Conservative 2; Mismatches 11; Indels 2; Gaps 2;

QY 26 PLPEANFPSPFLRH-HKHPLKPDNDQPPPOS 54  
 DB 4 PPDQNFYHHPHTHTHP-HPHSHPHPHS 32

RESULT 13  
 GDBX\_WHEAT STANDARD; PRT; 302 AA.

AC P21292;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 01-MAY-1991 (Rel. 18, Last annotation update)  
 DE Gamma-gliadin precursor.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Triticum.  
 ON NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. YAMHILL;  
 RA Scheets K., Hedgcock C.;  
 RT "Nucleotide sequence of a gamma gliadin gene: comparisons with other  
 RT gamma gliadin sequences show the structure of gamma gliadin genes and  
 RT the general primary structure of gamma gliadins.";  
 RL Plant Sci. 57:141-150(1988).  
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.  
 CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY  
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO  
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR  
 CC NEAR DIRECT REPEATS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M36999; AAA34272.1; -  
 DR PIR; JAO153; JAO153.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR001768; Cereal\_tryp\_amy1\_inh.  
 DR InterPro; IPR001956; Glia\_gluutenin.  
 DR InterPro; IPR000528; Plant\_LTP.  
 DR Pfam; PF00279; LTP; 1.  
 DR Pfam; PF00234; tryp\_alpha\_amy1; 1.  
 DR PRINTS; PR00208; GLIADGLUTEN.  
 DR SMART; SMO0499; AAI; 1.  
 KM Seed storage protein; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 19  
 FT CHAIN 20 302 GAMMA-GLIADIN.  
 FT SEQUENCE 302 AA; 34300 MW; 672437C770D6E782 CRC64;

Query Match 16.2%; Score 68.5; DB 1; Length 302;  
 Best Local Similarity 35.7%; Pred. No. 2.8;  
 Matches 20; Conservative 3; Mismatches 20; Indels 13; Gaps 3;

QY 26 PLPEANFPSPFLRH-HKHPLKPDNDQPPPOS 54  
 DB 93 PPDQNFYHHPHTHTHP-HPHSHPHPHS 32

RESULT 14  
 MAKE\_HUMAN STANDARD; PRT; 947 AA.

AC Q99558;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mitogen-activated protein kinase kinase 14 (EC 2.7.1.37) (NF-  
 DE kappa beta-inducing kinase) (Serine/threonine protein kinase NIK)  
 DE (HSN1K).  
 GN MAP3K14 OR NIK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-429 AND LYS-430.  
 RX MEDLINE=97172277; PubMed=9020361;  
 RA Mallin N.L., Boldin M.P., Kovalenko A.V., Wallach D.;  
 RT "MAPK-related kinase involved in NF-kappaB induction by TNF, CD95 and  
 RT IL-1.";  
 RL Nature 385:540-544(1997).  
 CC -1- FUNCTION: Lymphotoxin beta-activated kinase which seems to be  
 CC exclusively involved in the activation of NF-kappa-B and its  
 CC transcriptional activity. Induces the processing of NF-kappa-B  
 CC 2/P100. Could act in a receptor-selective manner (by similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphorylated  
 CC -1- SUBUNIT: Binds to TRAF2, TRAF5, TRAF6, IKKA AND NF-kappa-B 2/P100  
 CC (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Weakly expressed in testis, small intestine,  
 CC spleen, thymus, peripheral blood leukocytes, prostate, ovary and  
 CC colon.  
 CC -1- PTM: Autophosphorylated.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC MAP KINASE KINASE KINASE SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Y10256; CAA71306.1; -  
 DR MTM; 604655; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_AMP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_SF; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KM Phosphorylation.  
 FT DOMAIN 400 655 PROTEIN KINASE.  
 FT NF\_BIND 404 414 ATP (BY SIMILARITY).  
 FT BINDING 429 429 ATP (BY SIMILARITY).  
 FT ACT\_SITE 515 515 BY SIMILARITY.  
 FT MUTAGEN K->A: LOSS OF AUTOPHOSPHORYLATION; WHEN  
 FT SEQUENCE 947 AA; 104097 MW; D307F76993BC9661 CRC64;

Query Match 16.0%; Score 68; DB 1; Length 947;  
 Best Local Similarity 37.7%; Pred. No. 11;  
 Matches 23; Conservative 5; Mismatches 25; Indels 8; Gaps 3;

QY 4 PFHCRIGSYVRLRLKKEVLRPLPEANFPSPFLRH-HKHPLKPDNDQPPPOS 61  
 DB 232 PLDC-LNHVWKLNHPDGGRLPLRHPFYSRLRHPFPRHPL-OPKRPBLESFLG 285

QY 62 K 62  
 DB 286 K 286

RESULT 15

SALM\_DROME STANDARD: PRT: 1355 AA.

ID SALM\_DROME

AC P39770;

DT 01-FEB-1995 (Rel. 31, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Homeotic protein spalt-major.

GN SALM.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-94139659; PubMed-7905822;

RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,

RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaechle H., Schuh R.;

RT "Spalt encodes an evolutionarily conserved zinc finger protein of

RT novel structure which provides homeotic gene function in the head and

RT tail region of the Drosophila embryo.";

RL EMBL J. 13:168-179(1994).

CC -! FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST

CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY

CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE

CC TRANSCRIPTION OF THE TSH GENE.

CC -! SUBCELLULAR LOCATION: Nuclear (Probable).

CC -! DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND

CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS

CC WELL AS IN THE DEVELOPING TRACHEA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: X75541; CAA53229.1; -.

DR HSSP: P15822; 1BBO.

DR FlyBase: FBgn0004579; salm.

DR InterPro: IPR000822; znf-C2H2.

DR Pfam: PF00096; zf-C2H2; 7.

DR PRINTS: PR00048; ZINCINGER.

DR SMART: SM00355; znf\_C2H2; 7.

DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 7.

DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 7.

KW Developmental protein; zinc-finger; Metal-binding; DNA-binding;

KW Nuclear protein; Transcription regulation; Repeat.

FT ZN\_FING 451 473 C2H2-TYPE.

FT ZN\_FING 479 501 C2H2-TYPE.

FT ZN\_FING 824 846 C2H2-TYPE.

FT ZN\_FING 852 874 C2H2-TYPE.

FT ZN\_FING 884 906 C2H2-TYPE.

FT ZN\_FING 1289 1311 C2H2-TYPE.

FT ZN\_FING 1317 1339 C2H2-TYPE.

SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;

Query Match

16.0%; Score 68; DB 1; Length 1355;

Best Local Similarity 25.8%; Pred. No. 16;

Matches 23; Conservative 7; Mismatches 33; Indels 26; Gaps 5;

OY 1 GPRPFHCRQIGSYRLRLPKGEV-----LPLPEANPSPRLPHNK---HP----- 43

DB 475 GERPFKCNVCGSRF-----TTKGLKXVHFRHNAKPRHVMNATPRIHNDKFRPLDDM 530

OY 44 ----LKPDNQPFQSVSESCPGKFKSGFP 68

DB 531 SPTDSSPNHSPAPPLIG-SAPASFPAPFP 558

Search completed: May 24, 2002, 17:03:25

Job time: 860 sec

THIS PAGE BLANK (USPTO)